Social Comparison and Health-Related Judgement and Decision Making

by

Karen Lesley Melrose

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Psychology

University of Warwick, Department of Psychology
April 2016

Table of Contents

List of Tables

List of Figures

Acknowledgements

I am extremely grateful to so many people who have encouraged and supported me over the last four years, without whom this thesis would not have been possible.

First and foremost, I would like to thank my supervisor, Professor Gordon Brown, for his patience, guidance, wisdom, advice and feedback, not just over the course of my PhD but during my Master's degree as well. I really appreciate all the time that he has invested in my supervision and the help and support that he has given me over the years. It has been an honour working with him. I would also like to thank Professor Alex Wood for his feedback and advice and the ideas for the general health (Chapter 5) and social support (Chapter 6) studies. I would particularly like to thank him for helping me develop my writing skills during the time he spent reviewing and commenting on many drafts of the first manuscript that I wrote which described findings from my Master's research.

My thanks are extended to Dr. Nicole Tang for her help, knowledge and advice when planning the sleep study (Chapter 4) and to Professor Elizabeth Maylor for providing me with the opportunity to work on a research project with two of her third year undergraduate students. I would also like to thank Dr. Derrick Watson for being my mentor and for his support and guidance which was always greatly appreciated.

Thank you to the Behavioural Insights Team who provided me with the amazing opportunity to join them as an intern during my PhD. I learnt an incredible amount during my time with them which has undoubtedly made me a better researcher.

I am extremely lucky to have such wonderful friends and family and although space prevents me from doing so individually, I would like to thank them all for their encouragement and support, particularly over the course of my PhD. I do, however, especially want to thank my husband Ben who has been an incredible support to me during this time (and who inspired me to take this journey in the first place) and my Mom and Dad for all their love and support too. Special thanks also to my cat Minty who sat beside me whilst I wrote this thesis, providing both company and support.

Finally, I would like to thank the Economic and Social Research Council for funding my PhD and the Department of Psychology at the University of Warwick for their support and additional funding.

Declaration and Inclusion of Material from a Prior Thesis

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree apart from Table 3.4, which shows the results of research previously submitted for my Master's degree. This table is included for reference purposes only as I replicated the study during my PhD and discuss how the current results compare to the previous. The work presented (including data generated and data analysis) was carried out entirely by the author.

Inclusion of Published Work

Parts of this thesis have been published by the author. The aforementioned Table 3.4 was taken from a publication that describes my Master's research:

Melrose, K. L., Brown, G. D. A., & Wood, A.M. (2013). Am I abnormal? Relative rank and social norm effects in judgments of anxiety and depression symptoms severity. *Journal of Behavioral Decision Making, 26,* 174–184.

Chapter 6 includes the following publication:

Melrose, K. L., Brown, G. D. A., & Wood, A.M. (2015). When is received social support related to perceived support and well-being? When it is needed. *Personality and Individual Differences, 77,* 97–105.

Professor Gordon Brown and Professor Alex Wood contributed to the planning of this research and provided feedback on the questionnaires used in the studies and drafts of the manuscript.

List of Abbreviations

-2LL: -2 Log-likelihood

AIC: Akaike information criterion

ALT: Adaptation level theory

ANOVA: Analysis of variance

ASSIS-S: Arizona social support interview schedule - satisfaction subscale

BIC: Bayesian information criterion

DbS: Decision by sampling

GAD: Generalised anxiety disorder

MSPSS: Multidimensional scale of perceived social support

PHQ: Public health questionnaire

RFT: Range-frequency theory

SRH: Self-rated health

Abstract

The six studies presented in this thesis investigated the extent to which people compare to others when making a range of health-related judgements and decisions and aimed to identify the cognitive mechanisms used in this comparison process. A key question was whether biases found in specific judgements and decisions, such as deciding to seek help when it is not needed or not seeking help when it is needed, could be explained by social comparison effects. It was found that participants compared to others using rank-based strategies when making judgements and decisions about mental and physical health symptoms and when judging their health in general (Study 1, 2 and 4). Social comparison effects were generally small to medium in size (average Cohen's f^2 = 0.09, range = 0.01-0.39). Health-related help-seeking accuracy was associated with how participants believed their experience of symptoms compared to that of others. Participants were four times more likely to seek help when it was not needed if they believed that they experienced symptoms more frequently than others, and two to three times more likely not to seek help when it was needed if they believed that they experienced symptoms less frequently than others (Study 1). However, participants' beliefs about how their sleep compared to that of others had little influence on their sleep-related judgement and decision-making (Study 3). There was no evidence that participants' beliefs about how much support they received relative to others was associated with perceptions of this support (Study 5 and 6). The findings have implications for the development of both interventions that may improve accuracy in health help-seeking decisions and social norms-based interventions, the measurement of comparison effects, self-rated health and social support, social comparison theory, and models of symptom appraisal and health-related help-seeking.

Chapter 1: Introduction

Overview

Why do people make inaccurate judgements about their health? Why do people seek help for health problems when they do not need to or fail to seek help when they do? Why do people worry unnecessarily about their sleep? Why do people's perceptions of their social support not accurately reflect the amount of support they receive? The studies reported in this thesis attempt to resolve all of these questions with a single answer. When making judgements about their general health, their symptoms, their sleep, and their social support, and when making decisions about whether to seek help for health problems, people do not base their judgements and decisions on their absolute experiences alone, but on how these experiences compare to others. As will be explained, the use of social comparison for self-evaluation is ubiquitous in everyday life but can lead to inaccurate judgement and decision making if judgements and decisions are largely based on beliefs about how one compares to others. Such inaccuracy occurs when the distribution of experiences within the comparison sample is not representative of the actual distribution of the experience being judged.

For example, when judging how severe a cold is, it is hypothesised that people will take into consideration how their absolute experience of the symptoms (such as how long they have lasted for, the intensity of the symptoms, the number of symptoms they are experiencing) compares to other people's experiences of cold symptoms. If, for example, a person compares on the dimension of symptom duration and they have experienced cold symptoms for 7 days but the people that they are comparing to only experienced such symptoms for a few days before they felt better, they may be inclined to think their experience is "severe" when actually it is not (in fact the majority of people in the general population experience a cold for 7-10 days). This may have a knock-on effect to help-

seeking behaviours; for example, the above person may decide to go to the doctor about their cold after making the (inaccurate) judgement that their cold is “severe”.

The studies aim not only to examine whether such judgements are relative in nature but also to determine precisely *how* people compare to others (if they do). Cognitive models of judgement are applied to the domains of health, sleep, and social support for the first time in an attempt to understand the exact mechanisms underlying comparisons – a topic which has previously received little attention in the social comparison literature.

In understanding how people compare to others when making judgements about themselves, one can understand how people may make inaccurate judgements when comparison samples are not representative of the actual state of the world and also how such inaccurate judgements may be corrected. Correcting misperceptions about the self through correcting beliefs about others is the focus of norm recalibration interventions and the main implications of the research findings reported in this thesis lie in informing such intervention and information campaigns. This is explained in detail in the implications section of this Introduction. The research findings also have implications for the measurement of constructs such as social support and self-rated health. These and other implications relating to each domain studied are described in detail in the discussion sections of each chapter and in the General Discussion (Chapter 7).

The rest of this introduction takes the following format. First, a brief overview of social comparison and an explanation of how the current studies fit within this literature are given. Detailed reviews of the literature on each topic covered (i.e., physical and mental health help-seeking, general health, sleep and social support judgements), along with descriptions of how the current studies fit within their respective literature areas, are given at the start of each chapter. The summary on social comparison is followed by an overview of the cognitive models that will be tested in the studies. A summary of the main

motivation for and potential implications of the studies is then given followed by an outline of the thesis.

Social Comparison

"The human mind is a remarkable comparison processor. Whenever information is perceived, processed, or evaluated, it seems this information is compared to a salient context, norm, or standard." (Mussweiler & Epstude, 2009, p. 1)

We learn to compare stimuli at a very early age (Gentner & Medina, 1998; Gentner & Rattermann, 1991) and our use of comparison is pervasive in our everyday lives. As the quote above suggests, it has been proposed that when any kind of stimulus is processed, a frame of reference is evoked to which the stimulus is then compared (e.g., Kahneman & Miller, 1986). The use of comparison has been studied across many different areas in psychology and the idea that we compare to other people in order to evaluate ourselves and our behaviours has long been researched in the social literature (e.g., social comparison theory: Festinger, 1954; reference group theory: Hyman, 1942, Merton & Kitt, 1950; norm theory: Kahneman & Miller, 1986; relative deprivation theory: Runciman, 1966). This research, particularly work on social comparison theory, has highlighted the important role that social comparisons play in making judgements about ourselves and others and this research is briefly reviewed now.

Social Comparison Theory

Since Leon Festinger outlined the first systematic theory of social comparison in 1954, there has been much research on the reasons for social comparison, the choice of comparison target/direction of comparison and the affective and cognitive outcomes of comparison (for a detailed review of the history of social comparison theory see Suls & Wheeler, 2000). Traditional social comparison research has concentrated on investigating these questions in the context of the facilitation of three goals: self-assessment, self-

enhancement, and self-improvement (e.g., Festinger, 1954; Taylor & Lobel, 1989; Wills, 1981; Wood, 1989).

Festinger (1954) proposed that the goal of social comparison was self-evaluation of opinion and ability. He hypothesised that people have an innate desire to hold accurate assessments of their own and other people's opinions and abilities as such assessments are needed if an individual is to function effectively in their environment. When the preferred objective or physical standard is unavailable for evaluation, rather than facing uncertainty, people assess their own standing on a particular dimension by comparing to other people's standing on that dimension. In further development of the model, Schachter and Singer (1962) extended the use of social comparison in self-assessment to any situation under uncertainty or threat and not just the evaluation of opinions and abilities.

People also purposefully engage in social comparison in order to create, maintain, or confirm a positive self-image. To this end, they may make downward comparisons to people who are worse off than themselves using a contrastive process (Wills, 1981) or upward comparisons to those that are better but similar to them using an assimilation process (Wheeler, 1966; Collins, 1996). In both cases, social comparisons lead to increased self-esteem and positive affect. There is also research to suggest that, when self-esteem is threatened (particularly that of people from stigmatised groups), people make lateral, within-group comparisons. This is to avoid the loss of self-esteem that may result from comparing to others who may be worse off on a particular dimension but could be perceived as being better off in general because they belong to a more advantaged group (Crocker & Major, 1989; Major, 1994).

Finally, people compare to others for inspiration and/or motivation to improve themselves (Collins, 1996; Lockwood & Kunda, 1997; Taylor & Lobel, 1989). This occurs through a process of upwards comparison but only under certain conditions; the person comparing must have high self-esteem and must be comparing to others that are not seen

as competitors and whose success is perceived to be attainable (e.g., Taylor & Lobel, 1989). Of these three goals, the studies outlined in this thesis are concerned with the use of social comparison for self-assessment and the rest of this overview will focus on the research relating to this goal.

Comparison Sample Selection for Self-Assessment

Research on who people compare to when making social comparisons has suggested that comparison samples are selected based on the goals and motives of the comparison (see Biernat & Billings, 2001 or Corcoran & Mussweiler, 2010 for a review). There is a general consensus that, when the goal is self-assessment, people compare to similar others (e.g., Festinger, 1954; Kahneman & Miller, 1986; Suls, Gastorf, & Lawhom, 1978; Taylor, Neter, & Wayment, 1995; Wheeler, 1966). Accounts developed within traditional social comparison theory suggest that comparison sample selection is a complex and effortful process. For example, in the case of self-assessment, Goethals and Darley (1977) propose that people select others who are similar not only on the dimension that is being assessed but on attributes that may also influence one's standing on that dimension. This means that people have to first decide which attributes are relevant to the dimension being assessed and then construct a comparison sample of people who are similar on each of these attributes as well as on the dimension on which they are actually making the comparison (Goethals & Darley, 1997; Wood, 1989). Zanna, Goethals, and Hill (1975) give the example of a swimmer evaluating their speed. In order to do this the swimmer would consider the speed of other swimmers (the dimension being assessed) and dimensions that are related to swimming speed such as age, experience, and practice, and compare to people who are similar on all of these dimensions.

However, there is evidence to suggest that comparison processes can be performed effectively under cognitive load (Gilbert, Giesler, & Morris, 1995) and that social comparisons can occur spontaneously with the mere presence of others being enough to

influence self-assessment and behaviour (Morse & Gergen, 1970; Seta, 1982). These findings suggest that social comparison occurs with much less effort that originally thought and, in some cases, may even occur automatically as some studies have demonstrated how self-assessments have been affected by subliminal presentations of comparison targets (Mussweiler, Rüter, & Epstude, 2004). Furthermore, there is evidence to suggest that social comparison occurs whether objective standards are available or not (Klein, 1997). This has led to the suggestion that social comparison can be seen as a self-judgement heuristic (e.g., Corcoran & Mussweiler, 2010).

Social Comparison as a Heuristic

Mussweiler and colleagues have shown how comparative information processing can be quick and efficient (Mussweiler & Epstude, 2009) and can reduce uncertainty in judgements (Mussweiler & Posten, 2012). They propose that these are the reasons why comparisons are used so ubiquitously – i.e., they are being used as a heuristic. Mussweiler and Epstude (2009) argue that using objective standards to judge targets requires a lot of cognitive capacity as it may involve searching for information that is often difficult to obtain. They suggest that comparative information processing is more efficient than non-comparative processes because it 1) limits the search for knowledge about the target (information focus) and 2) substitutes knowledge about the target that is difficult to obtain or unavailable with easily accessible knowledge about similar previously encountered targets (information transfer). In a series of experiments, Mussweiler and Epstude (2009) showed that participants primed to think comparatively looked at fewer pieces of information about a target they were judging, showed evidence of information transfer about a similar, known target to a new, unfamiliar target they were judging, and made quicker judgements than control participants who were not primed. Mussweiler and Posten (2012) added to these findings by showing, in three experiments, how participants primed to think comparatively felt more certain about subsequent judgements than

controls. Although the targets being judged in both of these studies were not social ones, the authors conclude that it is likely that the same processes are involved with self-judgement through social comparison.

Underlying Mechanisms of Social Comparison

Whether people compare automatically, use social comparison to make quick self-judgements, or spend time constructing comparison groups in order to make self-assessments that are as accurate as possible, the underlying mechanisms of the comparison process (i.e., how people actually compare to others) have received little attention in the literature to date. Investigation of social comparison processes has largely focused on how the comparison leads to assimilation or contrast with others and the cognitive and affective outcomes of this (e.g., Markman & McMullen, 2003; Mussweiler, 2003). However, models of these processes do not explain what actually happens when the comparison is taking place.

The studies outlined in this thesis aim to investigate precisely this by applying cognitive models of judgement to a range of social comparisons. It is hypothesised that, when making self-judgements, people will construct a comparison sample (from memory and/or the immediate context) and then, through a series of binary, ordinal comparisons, calculate their rank within this sample and use their ranking as the basis for the judgement. For example, when judging the severity of a symptom a person might bring to mind a comparison sample of five other people they know who have also experienced the symptom. They will then judge how the severity of their experience of the symptom compares to the severity of other people's experiences. If, for example, they believe that four out of the five people had a less severe experience than they did then they are likely to conclude that their experience is quite severe as their experience ranks high within their comparison sample – second out of six. If, however, they believe that four out of the five people had a more severe experience of the symptom than they did then they are likely to

conclude that their experience was not that severe as the majority of others were worse off than them (i.e., the rank fifth out of six). This account is proposed by the decision by sampling model (Stewart, Chater, & Brown, 2006) and is explained, along with other competing models of relative judgement, in the next section.

Cognitive Models of Judgement

Throughout the studies in this thesis, I apply two competing models of relative judgement – decision by sampling (DbS) and adaptation level theory (ALT: Helson, 1947, 1948) – to self-judgements in the domains of health, sleep, and social support in order to investigate the extent to which, and how, people compare to others when making these judgements. These models are explained in the sections below, which give a brief overview of the history of relative models of judgement.

Adaptation Level Theory

ALT was originally developed to account for context effects in psychophysical judgements. Helson (1948) showed that participants did not judge the heaviness of a weight in absolute terms – their judgement of a target weight was influenced by the heaviness of a weight that they had held before judging the target (an anchor). Participants judged the target to be heavier when it was lifted after lifting an anchor weight that was lighter than the target than when they lifted it after lifting an anchor weight that was heavier than the target.

ALT proposes that when a target item is judged it is evaluated in comparison to previously encountered similar and contextually salient items. It is suggested that people have internalised reference levels ('adaptation levels') to which relevant new items are compared and that these reference levels are formed from past experience of similar and contextually salient items. This means that adaptation levels are constantly evolving with experience and are therefore different for everyone. In its simplest and commonly used form, ALT states that the judgement of new items in relation to the adaptation level

involves a comparison of the new item to the average of relevant previously encountered items. Helson's basic equation for the adaptation level (*AL*) is:

$$AL = K(M^pA^qR^r) \quad (1)$$

where *K* is an empirical constant, *M* is the geometric mean of previously encountered, similar items, *A* is the anchor item (if applicable), *R* refers to residual stimuli (later deemed irrelevant: Helson, 1964), and *p*, *q*, and *r* are constant weighting coefficients which sum to the value of 1 and are determined empirically.

ALT has been surpassed by rank and range-based models (described next) after theoretical and empirical limitations were uncovered (e.g., Birnbaum, 1974; Parducci, 1963; 1965; Sarris, 1967; 1971). However, it is still widely assumed in social psychology that we compare to the average of a comparison sample. For example, empirical studies testing predictors and consequences of social comparison often ask participants to evaluate themselves relative to an 'average' target. This method has been used often in health research, for example, in studies where patients have been asked how they think they compare to the 'average' patient with the same illness (see Arigo, Suls, & Smyth, 2014, for a review of such studies). Furthermore, studies implementing social norm-based interventions tend to give people information about how their behaviour compares to the average of the comparison group in an attempt to adjust their behaviour in line with this average. For example, in order to try to reduce energy consumption, Schultz, Nolan, Cialdini, Goldstein, and Griskevicius (2007) provided households with information about the amount of energy consumed by their household and the average amount of energy consumed by households in the neighbourhood. Similarly, in an attempt to increase participation in food waste recycling, Nomura, John, and Cotterill (2011) provided households with information about the percentage of households in their street that participated in food waste recycling and the average participation rate for the geographical area. As it is so often assumed that when people are asked to compare to others they

compare to the average, this assumption is formerly tested in the studies reported in this thesis through the application of ALT. This is done by investigating whether the difference between a person's experience and what they believe the average experience of others in the comparison sample to be predicts associated outcome measures. This is referred to as the participant's 'distance from the average'.

Range-Frequency Theory

Range-frequency theory (RFT: Parducci, 1965; 1995) was developed to account for findings that suggested that ALT could not adequately explain category judgement. For example, Parducci (1965) showed that participants sometimes made very different judgements about items when the adaptation level was held constant and made the same judgements about items when it was not. For example, in a series of studies, Parducci (1963, 1965) presented participants with 9 different sized squares multiple times in a block of 45 presentations. Participants were shown the whole block of presentations once and were then shown it again and asked to judge the size of each square on a scale from 1 = "very small" to 6 = "very large". Using a between-subjects design, Parducci changed the number of times each of the nine squares was presented in the block across groups of participants to manipulate the average size of all the squares presented (i.e., the adaptation level). He found that participants sometimes judged squares that were the same difference in size from the average square size of the block as being different in size. He also found that participants sometimes judged squares that did not have the same size difference compared to the average square size as being the same in size. This would not be expected if the participants were judging the size of each square by comparing it to the average size of all the squares in the set as ALT proposes.

Parducci (1965) proposed that people are influenced by the whole distribution of contextual items and not just the mean of these items. When a target (e.g., a person: x_i) is

evaluated within a context (e.g., a sample of people: x_1, x_2,…, x_n) the resulting judgment (J_i) is based on a compromise between two principles: range (R_i) and frequency (F_i).

$$J_i = wR_i + (1 - w)F_i \quad (2)$$

When the comparison sample is ordered on the dimension being judged [x_1, x_2,…, x_i,…, x_n] R_i is the target's position in relation to the highest and lowest people in the sample:

$$R_i = \frac{x_i - x_1}{x_n - x_1} \quad (3)$$

F_i is the target's relative ranked position within the ordered sample:

$$F_i = \frac{i - 1}{n - 1} \quad (4)$$

and w is a weighting constant (between 0 and 1) that reflects the influence of the two principles on the judgement.

RFT gained much empirical support not only in psychophysics where it originated (Parducci, 1963; Parducci & Perrett, 1971; Riskey, Parducci & Beauchamp, 1979) but in other, diverse domains and can account for context effects in judgements of body image (Wedell, Santoyo, & Pettibone, 2005), emotion (Rusell & Fehr, 1987), facial attractiveness (Wedell, Parducci, & Geiselmen, 1987), happiness (Wedell & Parducci, 1988; Smith, Diener, & Wedell, 1989), morality (Parducci, 1968), prices (Niedrich, Sharma, & Wedell, 2001), and psychopathology (Wedell, Parducci, & Lane, 1990). Many of these cases involve social comparison.

Decision by Sampling

DbS proposes that judgements of a target within a context depend solely on the target's relative ranked position within the context. It offers a process-level account of the frequency principle of RFT outlined in equation 4 above. Applied to self-judgements using social comparison, the model proposes that people will bring to mind a sample of others and will base their judgement about themselves on where they rank amongst these others on the dimension that is being judged. The model suggests a very undemanding

comparison procedure in which only two simple cognitive processes are used - ordinal comparison and frequency accumulation. A person's rank within their sample is their subjective self-evaluation and this is calculated by making a number of binary, ordinal comparisons to ascertain whether one is "better than", "equal to", or "worse than" each person within the sample and by keeping track of the number of comparison outcomes that either favour or do not favour oneself. The model therefore assumes that comparisons are ordinal in nature - consistent with research suggesting that people are better at discriminating between stimuli than they are at evaluating their magnitude (Miller, 1956; Stewart, Brown & Chater, 2005). It also assumes that people are able to encode, manipulate, and recall frequencies with relative ease - long assumed in humans and animals (see Sedlmeier & Betsch, 2002, for a review). Consistent with information sampling models of judgement (e.g., Fiedler, 2000; Juslin, Winman, & Hansson, 2007), comparison samples may be constructed from memory and/or the immediate context.

The model's simplicity lends itself to the idea that people may use social comparison as a heuristic to judge themselves but at the same time the model is also able to account for traditional views that comparison is a more effortful process whereby comparison samples are carefully constructed based on a number of different attributes. For example, people may make quick self-judgements if they are limited in time, capacity, or motivation by comparing to others that are easily retrieved from memory. Corcoran and Mussweiler (2010) refer to such people as "routine standards" and present evidence of their use. Or, they may base self-judgements on how they compare to retrieved implicit norms that they hold such as their beliefs about the distribution of the dimension that is being judged (for example, symptom severity) within the general population (Alicke, 1985; Krueger & Clement, 1994; Suls, 1986; Weinstein, 1980). On the other hand, when accuracy is important, time can be spent constructing the best comparison sample to compare to but the comparison process remains the same.

Considerable empirical support for the DbS model exists including evidence from a number of studies that have applied the model to social comparisons and have shown how an individual's rank within a comparison sample on a specific dimension predicts outcomes related to that dimension. These include attitudes towards the riskiness of alcohol consumption (Wood, Brown, & Maltby, 2012), concerns about indebtedness (Aldrovandi, Wood, Maltby, & Brown, 2015), judgements of the severity of crimes and punishments (Aldrovandi, Wood, & Brown, 2014), mental distress (Wood, Boyce, Moore, & Brown, 2012), perceptions of the health benefits of exercise (Maltby, Wood, Vlaev, Taylor, & Brown, 2012), student satisfaction (Brown, Wood, Ogden, & Maltby, 2015), and whether people think they have a mental health disorder (Melrose, Brown, & Wood, 2013). These studies utilised a novel methodology (explained in Chapter 2) that elicited participant's beliefs about the distribution of the dimension of interest (e.g., alcohol consumption, symptom severity, etc.) in a specified comparison sample (e.g., the general population). Previous research testing ALT and RFT has made assumptions that participants' beliefs about such distributions are correct. However, investigation into individual differences in participants' beliefs about others carried out in these studies has shown that this is often not the case. This will be further investigated in the current studies.

RFT vs DbS

The DbS model is used to investigate comparison effects over the RFT model for a number of reasons. Firstly, RFT only gives a descriptive account of rank and range effects whereas DbS offers a process-level account of rank effects. Therefore, as one of the main aims of the studies is to identify the processes underlying social comparisons, DbS is the more relevant model to use.

Secondly, although DbS does not predict effects of the range of a contextual distribution as RFT does, despite such effects often being observed; there is evidence to suggest that a purely rank-based approach such as DbS can account for these apparent

range effects. RFT has been tested mainly through presentation and manipulation of experimentally designed contexts for judgement in the laboratory. Testing of DbS on the other hand has focused more on remembered than experimentally constructed contexts; more reflective of actual, real-world judgement and decision-making (see references in the section above). This testing has shown that rank effects are present when participants use contexts retrieved from memory to make judgements and decisions, just as when experimental contexts were used when testing RFT. However, apparent range effects seen during RFT testing may actually be artefacts of using experimentally constructed distributions. In real-world judgement and decision-making, it would appear that range effects can be accounted for by a rank-based model when the distinctiveness of items in the context is considered. Based on the SIMPLE model of memory (Brown, Neath, & Chater, 2007), Brown and Matthews (2011) propose that when people construct a comparison sample of items by retrieving them from memory, the likelihood of a particular item being retrieved and therefore included is somewhat dependent on the items' discriminability, i.e., how distinctive it is relative to other relevant items. This means that when there are groups of many similar items that could be retrieved, the likelihood of them being so decreases, meaning that, overall, these items contribute less to the judgement being made. Brown and Matthews (2011) showed that there was no difference in the fit of a rank-based model that took into account the probability of items being included in the comparison sample based on their discriminability (which they refer to as the DbS-SIMPLE model) and the RFT model. They suggest that RFT should not be used over DbS in investigations of contexts effects just because it incorporates both range and rank effects (see also Brown et al., 2015, for an alternative account of how apparent range effects may reflect rank-based processes).

Finally, the studies reported in this thesis aimed to investigate how participant's beliefs about the world affected judgements and decisions they made about themselves.

This means that the context for judgement was not experimentally manipulated; it was retrieved by each participant. Therefore, DbS is again the more relevant model to use to investigate these kinds of context effects because, as previously mentioned, it focuses on comparison contexts retrieved from memory.

Given all of the above, there is little reason to test the RFT model over the DbS model in the current studies. Therefore, DbS and ALT accounts of judgements are compared directly in each study in order to investigate 1) the extent to which various judgements are made using social comparison and 2) precisely how people compare to others, i.e., whether they use rank-based strategies or their distance from the mean of the comparison sample. It is hypothesised that, when comparing to others to make judgements about themselves, people will do so using rank-based strategies (consistent with DbS). Understanding precisely how people make judgements about themselves through social comparison is useful for the development of interventions that aim to correct inaccurate self-judgements (and ultimately change behaviour) through correcting inaccurate beliefs about others. This will be discussed next.

Main Research Implications

Research has shown that the use of social comparison in self-assessment can lead to inaccurate judgements being made; these judgements may in turn affect behaviour and have adverse consequences. The classic example is alcohol consumption in university students. Students who drink a lot of alcohol typically perceive that their consumption is more 'normal' than it really is because they incorrectly believe that their peers are drinking more than they are. They overestimate consumption in others, which makes their consumption seem less extreme than it is (McAlaney & McMahon, 2007; Perkins, 2007; Perkins & Wechsler, 1996). Students come to incorrect judgements about themselves because they compare to others to make the judgement and their beliefs about others are incorrect.

Interventions have been developed with the aim of changing behaviour by changing people's perceptions of themselves through correcting their misperceptions of others. These norm recalibration interventions typically involve giving people information about descriptive social norms (how others usually behave) in one of two ways: through social marketing of the normative information using mass communication methods such as posters and flyers, or by using personalised feedback which provides information about how the individual's behaviour compares to the actual norm. These interventions have been found to be somewhat successful in changing a variety of behaviours, for example, decreasing alcohol (e.g., Agostinelli, Brown, & Miller, 1995) and energy consumption (e.g., Dolan & Metcalf, 2013; Schultz et al., 2007) and increasing recycling participation (Schultz, 1999; Nomura et al., 2011) and the amount of material recycled (Schultz, 1999).

However, these interventions tend to supply people with information about how they compare to the average of the comparison sample. They assume (without actually testing) that people compare their behaviour to the group average and that they are likely to modify their behaviour in line with this average (e.g., Dolan & Metcalf, 2013; Harries, Rettie, Studley, Burchell, & Chambers, 2013; Nomura et al., 2011; Schultz, 1999; Schultz et al., 2007). However, recent research directly comparing DbS and ALT has shown that when health-related judgements are made through a comparison to others they are made not through a comparison to the group average but through a comparison based on the individual's rank within the group. For example, Wood, Brown et al. (2012) showed how the rank of an individual's alcohol consumption amongst that of others predicted judgements relating to the risks of developing alcohol-related disorders. In this study, no evidence was found that individuals compared their drinking to that of the average of others in the comparison group in order to make the judgements. Similarly, Melrose et al. (2013) showed how participants' judgements of whether or not they thought they had depression or an anxiety disorder were predicted by where the occurrence of their

symptoms of these disorders ranked amongst others. These judgements were not predicted by the distance of participants' experiences of the symptoms from the average experience of the comparison group.

It is therefore possible that norm-based interventions and education campaigns may be more effective when rank-based information, rather than information on how people differ from the average, is supplied. The studies in this thesis therefore examine the nature of the social comparison process in detail. Furthermore, they are all carried out in areas where there is expected to be discordance between subjective and objective experiences that may lead to inaccurate social comparisons and hence may be targeted effectively by norm recalibration interventions. The extent to which people differ in their beliefs about the distribution of absolute experiences (such as symptom occurrence) in a comparison sample that is common to all participants (the general population) is explored throughout the thesis in an attempt to identify areas where inaccurate social comparisons may be particularly prevalent and therefore useful targets for such interventions.

Thesis Outline

This thesis is structured in the following way. First an outline of general procedures used in all of the studies is given in the next chapter (Chapter 2; on methodology). This chapter includes overviews of participant recruitment, procedures followed by participants, construction of questionnaires, methods used to investigate social comparison and compare ALT and DbS accounts, and data analysis. The four chapters that follow (Chapters 3-6) outline the six studies completed during my PhD. These each begin with a review of the literature relevant to each topic (physical and mental health help-seeking, sleep, general health, and social support) and an explanation of how the current studies fit within this literature. The studies are then explained, results provided and the implications of the findings are discussed in relation to each specific domain. General conclusions and implications are discussed in Chapter 7.

Chapter 2: Methodology

Introduction

This chapter gives an overview of how participants were recruited and compensated for all of the studies outlined in this thesis, the methods used to investigate whether and how people compare to others when making specific judgements and the general procedures used in all of the studies. There are six studies reported in this thesis that are fully explained in the forthcoming chapters. For the purpose of this chapter, the studies are referred to by name only and often in the order that they were undertaken which differs from the order in which they are presented in this thesis. The studies are listed below in the order they were completed along with their corresponding thesis chapter.

1. Social Support: Study 1 – Chapter 6
2. Social Support: Study 2 – Chapter 6
3. General Health – Chapter 5
4. Mental Health – Chapter 3
5. Physical Health – Chapter 3
6. Sleep – Chapter 4

Participants

Sample Size Calculations

The aim of the two social support studies was to investigate whether perceptions of support (such as satisfaction with received support) were better predicted by how an individual thinks the amount of support they receive compares to that received by others (i.e., whether they get more or less support than others) than simply by the actual amount of support that they receive. It was hypothesised that the relationship between received support and perceptions of support would be weaker than the relationship between how the received support compares to others and perceptions of support and sample size

calculations for these studies were based on detecting the weaker relationship. A previous meta-analysis identified the average correlation between received support and perceptions of support to be *r* = .35, indicating a medium sized effect (Cohen, 1988). Sample size calculations suggested that at least 49 participants would be needed to have an 80% chance of detecting a correlation of .35, and that, for multiple regressions with 5 predictor variables, at least 91 participants would be needed to have an 80% chance of detecting a medium sized effect. In Study 1, 198 participants were recruited and in Study 2 there were 202 participants. High numbers of participants were recruited to allow for any exclusions due to difficulties understanding the methodology used (please see the section below entitled 'measuring how an individual compares to others' for more details of this). The minimum number of participants included in the analyses across the studies was 128.

The remaining studies were slightly different in that their main aim was to investigate whether individuals compare to others when making specific health-related judgements and, if so, the extent to which they do so and how. Previous work investigating the effects of social comparison on health-related judgements that used a similar methodology to the present studies (Melrose et al., 2013; Wood et al., 2012) was used as a guide to calculating sample sizes for these studies. Both previous studies suggested that the kinds of social comparison effects that the current studies are investigating might be small. For example, Melrose et al. (2013) report that the relative rank of participants' experiences of depression and anxiety symptoms within the general population accounted for 6.2% and 4.3% of the variance in their judgements of whether they thought they had depression or anxiety respectively. These figures correspond to Cohen's f^2 values of 0.066 and 0.045 respectively; values of 0.02, 0.15, and 0.35 indicate small, medium, and large effects respectively. Sample size calculations for these studies were based on the anticipated main analyses (multiple regression with five predictor variables) and suggested that at least 643 participants would be needed to have an 80% chance of detecting a small sized effect. This

sample size was obtained for three out of the four studies: general health: *N* = 643, mental health: *N* = 643, and sleep: *N* = 656, but due to difficulties in collecting data for the physical health study (see below), a sample size of only 543 participants was collected for that study. This is still a large enough sample size to have a 75% chance of detecting a small effect in a regression analysis with 4 predictor variables, which was enough to test the main hypotheses (see Chapter 3). A different methodology was utilised in these studies which meant that recruiting higher numbers of participants to allow for exclusions was not required (again, please see the section below entitled 'measuring how an individual compares to others' for more details of this).

Recruitment

Due to the large numbers of participants needed, participants were recruited using two online crowdsourcing platforms: Amazon Mechanical Turk (MTurk) and CrowdFlower. On both these platforms, created tasks (in the case of all the studies in this thesis this was to complete a questionnaire) are made available to "workers" – people who sign up to the platform in order to complete tasks in exchange for money or Amazon vouchers in the case of MTurk. It is possible to set eligibility criteria on both platforms so that specific people can be targeted, for example, those living in a certain country, those above a certain age or those classified as "high quality" workers. There has been much research on the use of MTurk workers in psychological studies (reviewed briefly below) and this platform was chosen in preference to other crowdsourcing websites for this reason. However, due to the U.S. Patriot Act it is not possible to post tasks on MTurk unless you are a U.S. citizen and so an intermediary company, CrowdFlower, was used to create and post tasks on MTurk.

This worked well for the first three studies (social support 1 and 2 and general health) but then CrowdFlower stopped posting tasks on MTurk and started using different crowdsourcing channels. As no other way of accessing MTurk workers was available at that time, the next two studies to be carried out (mental health and physical health) were run

on CrowdFlower through these channels. There were no problems collecting the mental health study data[1], however, after this study was completed, CrowdFlower changed their practices and increased the number of crowdsourcing channels that they used substantially to over 100. When data collection for the physical health study started it quickly became apparent that these changes had resulted in a distinct reduction in the quality of participants. Many people rushed through the survey to get the payment code at the end and then used the code multiple times from different accounts (it was not possible to supply each participant with a unique payment code) in order to receive several payments. These problems had not been encountered in any of the other studies. The changes in practice meant that it was also possible for people to take the survey multiple times (despite blocking of multiple responses from the same IP address). This meant that it took much longer (over 3 months) to collect the data, as many participants had to be excluded. Fortunately, during this time, another company (MTurk Data) was found to run the final study (sleep) on MTurk and no problems were encountered during data collection for this study.

Research has shown that MTurk workers produce high quality data in psychological experiments (Buhrmester, Kwang, & Gosling, 2011) and reliable data in the area of judgment and decision making specifically (Paolacci, Chandler, & Ipeirotis, 2010). MTurk workers differ somewhat from the U.S. general population in their demographic characteristics. They are young – on average in their thirties, overeducated, underemployed and under-representative of Blacks and Hispanics whilst over-representative of Asians (Berinsky, Huber, & Lenz, 2012; Paolacci et al., 2010; Shapiro, Chandler, & Mueller, 2013). Despite this, they are more representative of the U.S. population than the university undergraduates typically used in psychological research as

[1] Although it took quite a bit longer to collect the data; a month as opposed to a few days to a week for the first three studies.

well as other internet samples in general (Paolacci et al., 2010). There is also evidence to suggest that MTurk workers are more attentive to instructions and questions in online surveys than university undergraduates (Hauser & Schwarz, 2015) and that they are truthful when providing self-report information (Rand, 2012).

Participants were paid $0.75 to $1.50 for completing questionnaires in the studies outlined in this thesis (which never took longer than 15-20 minutes to complete). This is above the average MTurk payment of around $1.40 per hour (Horton & Chilton, 2010). Compensation amount has been shown to have no effect on data quality in tasks that require subjective responses such as those that participants completed in the thesis studies (Buhrmester et al., 2011; Marge, Banerjee, & Rudnicky, 2010; Mason & Watts, 2009).

When tasks were posted on MTurk or CrowdFlower, they all had the same general instructions that were tailored to the theme of the questionnaire. These instructions contained a short overview of the task, for example: “You will be asked to complete a 15 minute online questionnaire about your health and the health of others. Please answer all questions; you may only complete the study once. You will receive $0.75 for your participation. Thank you for your honest answers in this task.” followed by the process that the participant needed to follow in order to complete the survey and be paid. Tasks were typically titled “Questionnaire on [topic] for the University of Warwick, UK”.

Questionnaire Design and General Procedure

As previously mentioned, in all of the studies data were collected using questionnaires. For the first two studies (social support 1 and 2) questionnaires were designed and hosted using Survey Gizmo and for the remaining studies Qualtrics was used as it offered more sophisticated question options (such as slider scales) and was able to randomise the presentation of question blocks for more effective counterbalancing. When participants clicked on the link to start the survey they first saw an instruction page that

contained the same instructions for each study but tailored to the topic of the questionnaire. For example:

“Thank you for agreeing to take part in this study which is investigating people's opinions about their general health and the health of others.

This study is being conducted by Karen Melrose (k.l.melrose@warwick.ac.uk), a PhD student in the Department of Psychology at the University of Warwick, UK, as part of her doctoral thesis. Should you have any complaints related to this study please contact the University of Warwick's complaints committee (details can be found at http://www2.warwick.ac.uk/services/rss/researchgovernance_ethics/complaints_procedure).

All data collected are confidential and data will primarily be used for the thesis and possibly in a research publication.

You will be asked to answer a series of questions that should take around 15 minutes to complete.

Please read the questions carefully and note that there are no right or wrong answers - we are simply interested in your thoughts.

Please complete the survey in one session and answer all questions.

You have the right to withdraw from this study at any time by closing the browser window and are under no obligation to continue with the study once you have started.

Please click on the 'next' button to give your consent to take part in this study and to start the survey.”

Participants then completed the survey questions (outlined in each chapter) and at the end of the survey always saw a debrief screen which contained a code for payment, a brief explanation of the study and the investigator’s e-mail address in case the participants wanted any further information.

Measuring How an Individual Compares to Others

As outlined in Chapter 1, one of the central hypotheses that are being tested throughout this thesis is that, when comparing to others in order to make judgements, people use rank-based strategies rather than comparing to the average of their comparison sample.

Standard Methodologies

Previous studies (e.g., Aldrovandi et al., 2013; 2014; Maltby et al., 2012; Melrose et al., 2013; Wood et al., 2012) have tested the above hypothesis in different contexts using a method referred to here as the 'distribution elicitation' method. Using this method participants are asked about their experience of the topic of interest, for example, in the case of Wood et al. (2012), how much alcohol they drink, and are also asked 11 questions eliciting what they believe to be the distribution of alcohol consumption amongst other people in a given comparison group (often the general population). These questions typically take the following form (this example taken from Wood et al., 2012): "The top *x*% of the UK adult population consume more than ____ units of alcohol per week on average?", where *x* typically takes values of 1, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 99 and participants are asked to fill in the blank. Answers to these questions are then used to estimate each participant's normal cumulative distribution function of their experience (e.g., alcohol consumption) within the general adult population. From this, the relative ranked position of a participant's experience within their elicited distribution can be calculated along with the distance of their experience from the mean of the distribution. There are however two problems with using this methodology. The first is that participants often find the distribution elicitation questions hard to understand and this results in many participants (typically around 30%) having to be excluded from the study. The most common reason for exclusion is that participants answer the questions with values going from low to high instead of high to low. For example, using the task outlined above, the top 1% of the population means the 1% of the population who drink the most so, naturally, the answer to this question should be the highest value of alcohol units answered. The values that participants give to subsequent questions should then decrease as the percentile increases – the top 20% of drinkers drink less than the top 1% but more than the top 30% etc. Quite often participants misinterpret the question, start low and then increase their

answers through the question set. In previous studies appropriate responding has been tested by calculating Kendall nonparametric correlation coefficients for each participant to evaluate the ranked correlation between their responses to the first question (should be high) to their last (should be low) and the values 11 to 1. Participants are then excluded if $\tau < .80$. The second problem is that, although based on participants' beliefs about the world, this method is an indirect way of measuring participants' beliefs about how they compare to others.

New Methodology

Therefore, during the course of my PhD, I decided to develop a new, direct method of measuring how a person believes they rank in comparison to others and how much they differ from the average of the comparison group. The new method involves directly asking participants 1) where they think they rank in comparison to others and 2) what they think the average experience of others is. Each participant's answer to the latter question is then subtracted from their own experience (always asked about) to give a measure of how much they believe they differ from this average. For example, in the sleep study participants were asked: "Out of 100 people, how many do you think have had more hours of actual sleep than you, on average, over the last 90 nights?" (rank measure) and: "What do you think the average number of hours of sleep that people had each night over the last 90 nights was?" (average measure). In this task, participants were asked to consider 'people' to be adults of about their age in the U.S. general population. As an aside, the general population is used as the comparison group in all of the studies (apart from general health when the focus of the study was to investigate what comparison groups people use) because it is a comparison group that is the same for all participants thus allowing for differences in beliefs about others to be investigated. In some studies participants were asked to think about people about their age in the general population and this was

because, in these instances, age-related differences in the topic that participants were being asked about, for example sleep and physical health, have been shown to exist.

The new rank and average measures were used in the mental health, physical health, general health, and sleep studies (the distribution elicitation method was used in both social support studies) and their inception was largely due to the design of these studies. As explained earlier, the size of the sample required for each of these studies was considerably greater than the sample required for the social support studies. With the new measures, all participants could be included in the analysis, which meant that no extra participants would need to be recruited to allow for inevitable exclusions due to inaccuracy in answering the distribution elicitation questions. This was useful as, due to financial constraints, it would not have been possible to carry out all of the studies if an extra 200 people per study had to be recruited, based on typical exclusion rates of around 30%. More importantly, use of the new measure meant that multiple independent variables could be measured with relative ease to the participants as they only had to answer one rank and one average question for each variable as opposed to a set of eleven questions. This allowed for a more thorough investigation of effects in the general health, physical health, and sleep studies. For example, in the general health study it was possible to ask participants about how they thought they compared to people in nine different comparison groups. In the physical health study, participants were asked about four symptoms they had experienced, and, in the sleep study, they were asked about five aspects of their sleep. This would not have been possible using the distribution elicitation method without making the questionnaire somewhat long and tedious. Therefore, for these three studies only the direct methodology was used in order to keep the questionnaires as short as possible in order to encourage participant engagement.

With regards to the mental health study, participants were only asked about their experience of two symptoms and so it was possible to include both methodologies (the

indirect distribution elicitation questions and the two new, direct rank and average measures) in the questionnaire in order to compare them, without the resulting experiment being too long. Correlations between participants' two different rank and distance from the average values resulting from the two different methodologies indicated good construct validity: Depression rank variables: r = .843, anxiety rank variables: r = .814, depression distance from the average variables: r = .748, and anxiety distance from the average variables: r = .748, all p values < .001.

Data Analysis

All studies utilised a correlational design and therefore correlation and regression analyses were used to test relationships between variables. Where group differences were investigated, chi-square, analysis of variance (ANOVA) and t-tests were also used. All analysis was conducted using SPSS.

Data Screening

Prior to analysis, data were screened for outliers using scatterplots. As all of the questions in the survey were mandatory, (participants were informed of this before they took part) there was never any missing data. The design of the survey meant that outliers could be controlled for to some extent in the majority of the studies by setting limits on the possible answers the participants could give. This was only ever done when there were natural limits, for example, when asking on how many of the last 90 days the participants felt tired during the day there is a natural maximum answer of 90. However, when asking about how many times a person has received support in the last month there is no natural limit so these questions were left without any answering bounds. In the studies where outliers were removed (physical health and sleep), how this was dealt with is explained in the relevant chapters. Assumption testing (e.g., linearity, normality, etc.) for the statistical test being used in the analysis was carried out after the initial data screening.

Hypothesis Testing

As previously mentioned, correlation and regression analyses were used predominantly throughout this thesis. Correlations were mainly undertaken to check for collinearity between independent variables before carrying out the regression analyses. Where correlations between independent and dependent variables are investigated and reported, Spearman's correlation coefficients are reported for correlations between continuous independent variables and ordinal dependent variables and Pearson's correlation coefficients are reported for correlations between continuous independent and dependent variables.

Regression analyses were tailored to the dependent variable and therefore binary logistic, ordinal logistic and multiple linear (ordinary least squares) regressions were used to investigate how well dependent variables were predicted by the independent variables. When conducting ordinal regression, the polytomous universal model (PLUM) and the link function that provided the best fit of the full model compared to the intercept-only model (i.e., the one that produced the biggest difference between the two -2 log likelihood values) was used. Regressions were always performed using the enter method. For consistency, bootstrapped confidence intervals, standard errors, and significance values of each predictor are reported throughout although assumptions of homoscedasticity and/or normally distributed residuals were violated only in some cases.

The main analysis described in each chapter involves testing the central hypotheses outlined in Chapter 1 - that people compare to others when making specific judgements and that they do so using rank-based strategies. This is primarily examined through regression analyses. Regression models reported always contain age and gender as covariates as, in the majority of cases, changes in the outcome variable across ages and genders have been shown and so these are controlled for in the model. Gender is always input as females = 0 and males = 1. The participant's absolute experience of the construct

being tested (e.g., days they have experienced a symptom, number of hours of sleep they have had) is also always included in the model as it is hypothesised that this will always have a bearing on the judgement being made. In many cases, previous research has assumed that judgements are made based on a person's absolute experience and one of the central arguments here is that people do not base judgements on their absolute experience alone but on how this experience compares to other people's experiences. Therefore, it is useful to be able to compare the effect size of absolute and comparison variables in the model. For linear regression models this was done by plotting standardised beta coefficients and their corresponding 95% confidence intervals (estimated via bias corrected bootstrap, 1000 re-samples) to determine whether the confidence intervals overlapped by less than 50% thus indicating a significant difference between the two beta weights (Cumming, 2009). Menard (2004; 2011) notes that, unlike in ordinary least squares regression, there is no single, widely accepted definition for a standardised coefficient in logistic regression and standardised coefficients are not computed by most statistical software. Menard (2004; 2011) outlines six different methodologies for calculating standardised logistic regression coefficients that can be used to ascertain the rank order of the importance of each predictor in the model but not significant differences between coefficients. Therefore, for the binary regressions one of these methods (multiplying the unstandardized coefficient of each predictor by the sample standard deviation of that predictor: Agresti, 1996; Menard, 1995) has been used to calculate semi-standardised coefficients which are reported in the relevant results tables and are used solely to understand the hierarchy of the importance of predictors in the models. When ordinal regressions are reported in the thesis, the same analysis is also reported using multiple linear regression to assess the strength of predictors in the model.

As it is hypothesised that people will either use rank-based comparison processes or will compare to the average of the comparison group when making judgements, the following regression models are presented in each chapter:

- Step 1: includes age, gender and the absolute experience,
- Step 2a: includes age, gender, the absolute experience and the rank of this experience within a comparison sample,
- Step 2b: includes age, gender, the absolute experience and the distance of this experience from the average experience of the comparison sample[2].

Evidence for the use of either rank or average-based social comparison is obtained from the variable's performance in the model and from comparing the fit of the step 2a and 2b models (non-nested models). For multiple linear regression models, this is done using Hotelling's t-tests to ascertain whether the correlations between the predicted and actual values of the outcome variable across the two models differ significantly. For binary and ordinal regression models, this can be done by calculating Akaike's information criterion (AIC) and/or Bayesian information criterion (BIC) for both models and assessing the difference between them – lower values indicate a better fit. AIC and BIC are calculated using the equations below:

$$BIC = -2 \ln L + k \ln N \quad (5)$$

$$AIC = -2 \ln L + 2k \quad (6)$$

Where L is the likelihood of the model given the data, k is the number of parameters (or, in this case, predictors) in the model and N is the sample size. However, as the step 2a and 2b models always have the same number of predictors (4) and the same sample size, the difference in AIC and BIC values for the models always equates to the difference in -2 log-

[2] Models which included all five independent variables (age, gender, absolute experience, rank and distance from the average) were also formulated. These are not presented in the thesis as it is assumed that people either use rank-based strategies or compare to the average, not both. The pattern of results seen did not differ between the two methods.

likelihood values. Therefore, the difference in -2 log-likelihood values is used to assess which model fits the data better. Differences in AICs and BICs are interpreted similarly; for both criteria a difference of less than or equal to two indicates hardly any difference between the fit of the models and a difference greater than ten indicates "very strong" evidence against the model with the higher AIC or BIC value (Burnham & Anderson, 2004; Kass & Raftery, 1995). Between these values, support against the model with the higher AIC or BIC value increases. BIC cut off points of differences between two and six indicating "positive support" against and differences between 6 and 10 indicating "strong support" against the model with the higher BIC (Kass & Raftery, 1995) are used to interpret the differences between the -2 log-likelihood values of the step 2a and 2b models.

Mediation analyses are used in Chapters 3 and 4 to test for indirect associations between social comparison and help-seeking behaviours that may be present based on previous findings. Indirect effects were tested using a bootstrap estimation approach with 1000 samples using the PROCESS custom dialog box for SPSS (Hayes, 2012). PROCESS also calculates the direct effects between the independent variable and mediator and the outcome variable and mediator.

Due to the large number of regression analyses undertaken for each of the chapters, the written account of the findings reported in each results section focuses only on the evidence for or against the core hypotheses presented. Full details of all of the statistics calculated are presented in tables. Although a large number of tests were undertaken in each of the studies reported, these were all pre-planned, deemed necessary to answer the research questions outlined, and, apart from in the social support studies, investigated different constructs. Therefore, it was decided that due to the use of multiple analyses of the same construct in the social support studies, a conservative alpha level of .01 would be used for all statistical tests in these studies and an alpha level of .05 would be used in all of the other studies.

Terminology

To recap, the main analyses presented throughout this thesis investigate whether three main variables predict related outcomes. The first of these is participants' actual experience of the construct of interest, for example, the number of days they experience symptoms, the number of hours sleep they get, and the frequency with which they receive support. This is often referred to as their "absolute experience". The other two variables are measures of how participants believe their actual experience compares to others. These "others" are always either people in the general population or people around the participant's age in the general population. Participants are asked to compare to these comparison groups in particular as they are shared between all or groups of participants. This means that the accuracy of participants' beliefs about others can be investigated. The first comparison variable measures where participants think their absolute experience ranks in comparison to others. This is referred to as the "rank" variable. For example, if a participant believes that only 30% of people in the general population feel tired on more days a month than they do then their rank is 70 out of 100. The second comparison variable measures how different the participant's absolute experience is from what they believe the average experience of others to be. This is referred to as the "distance from the average" variable. For example, if a participant feels tired 5 days a month but they believe that, on average, people in the general population feel tired on 10 days a month then their distance from the average is -5 days.

Ethical Approval

All studies were approved by the Department of Psychology Research Ethics Committee at the University of Warwick.

Chapter 3: Physical and Mental Health

Overview

The studies outlined in this chapter aimed to replicate and extend the findings of Melrose et al. (2013). That study, completed during my Master's degree, showed that participants' judgements of whether they thought they suffered from depression or an anxiety disorder were predicted by the ranked position of their depression and anxiety symptom occurrence (the number of days a month they felt depressed or anxious) within what they believed the distribution of symptom occurrence in the general population to be. This was over and above their absolute experience of the symptom and the distance of their experience from the mean of the distribution. The present studies aim to replicate these findings (Study 1) and to see whether participants make judgements about their physical health as they do their mental health (Study 2). However, the main focus of both studies is to see whether and, (if so) how, social comparison affects help-seeking behaviour for physical and mental health problems. This will be investigated directly and indirectly through perceptions known to influence help-seeking behaviour such as judgements about symptom severity and worry about symptoms. Therefore, this introduction focuses on explaining previous research on health help-seeking, the role of social comparison within this and how the current studies build on previous findings.

Introduction

There has been much research investigating the processes involved and factors affecting health help-seeking behaviours, predominantly in an attempt to understand why people delay or do not seek help when they need to or do seek help when they do not need to. Both instances of inaccurate help-seeking behaviour are problematic and prevalent although research has tended to focus on the former, perhaps because delaying or not seeking treatment has more serious consequences than seeking help when it is not

needed. For example, delay in seeking treatment for cancer or a heart attack can result in life-threatening consequences as survival rates decline with the advancement of symptoms over time. Despite this, research has shown that people do delay in seeking help for cancer and heart attacks even when they are experiencing indicative symptoms. For example, in a survey of 1500 women with ovarian cancer, 70% stated that they had been experiencing symptoms for 3 months before seeking help and 15% for more than a year (Goff et al., 1998). It has been estimated that around a third of women diagnosed with self-discovered breast cancer delay seeking help for at least 3 months with as much as a quarter delaying 6 months (Facione, 1993; Richards, Westcombe, Love, Littlejohns, & Ramirez, 1999). Other studies have shown that people experiencing heart attacks often wait hours before seeking care. Goff, Nichaman, Ramsey, Meyer, & Labarthe (1995) found that 22% of patients surveyed waited at least 6 hours after the onset of symptoms to seek help. Hedges et al. (1998) found the median time taken by heart attack patients surveyed from first experiencing symptoms or sensations to calling emergency services for help was 7.25 hours. Delay in treatment seeking for physical health problems is not specific to these disorders either. Ingham and Miller (1979) conducted a comprehensive medical screening of over 3000 people in one UK borough and found that 57% had a variety of symptoms warranting treatment that had never been sought. Scambler and Scambler (1985) refer to this as the 'illness iceberg'.

Delaying or not seeking help when it is needed is not just an issue in physical illness but in mental health as well. Evidence from general population surveys (e.g., Andrews, Issakidis, & Carter, 2001; Kessler et al., 2005; Mojtabai, Olfson, & Mechanic, 2002; Wang et al., 2005) suggests that around 30-40% of people with symptoms of mental illness have not sought help (Henshaw & Freedman-Doan, 2009). Untreated mental illness can be as costly as physical illnesses such as diabetes and heart disease (Druss, Rosenheck, & Sledge, 2000; Katon et al., 2008) in terms of the costs incurred from loss of productivity and wages in the

workplace (Adler et al., 2006), the impact mental illness has on physical illnesses such as hypertension (Katon & Ciechanowski, 2002) and the resulting overuse of primary care services for other reasons (Katon, 2003; White et al., 2008).

Although many people do not seek help for mental health problems when they need to, Mojtabai (2008) notes that often the people who do seek help experience only minor symptoms, distress, and impairment in daily functioning. It has been estimated that 20-40% of people attending primary care do not have a major illness and that in 30-60% of cases people present with symptoms that have no serious underlying cause (Backett, Heady, & Evans, 1954; Barsky, 1981; Kroenke & Mangelsdorff, 1989). Unnecessary use of health services is common for particular symptoms such as cold and flu symptoms (Braun et al., 2000). Such overuse wastes time and financial resources (be it the patient's or the government's) and may result in unnecessary exposure to iatrogenic diseases or complications (Peters, Stanley, Rose, & Salmon, 1998).

Why do people seek help for health problems when they do not need to and fail to seek help when they do? Stage process models (e.g., Andersen, Cacioppo, & Roberts, 1995; Cauce et al., 2002; Safer, Tharps, Jackson, & Leventhal, 1979; Shaw, Brittain, Tansey, & Williams, 2008) view physical and mental health help-seeking as a decision making process in which people first have to decide whether they are ill, then whether they require medical attention, and who the correct person or service is to receive this from. Decisions about help-seeking appear to be heavily based on the experience of abnormal somatic or psychological sensations – the vast majority of people that seek medical help report experiencing some kind of symptom or symptoms (Berkanovic, Telesky, & Reader, 1981; Cameron, Leventhal, & Leventhal, 1993; Costa & McCrae, 1980; Leventhal, Hansell, Diefenback, Leventhal, & Glass, 1996; Stoller, 1997). Therefore, the decision about whether or not you are ill (and the subsequent decision on whether or not to seek help) is likely to reflect the appraisal of detected symptoms which may take into account objective

information about the symptoms such as their duration and frequency, and perceptions about their severity, significance, intensity and cause (Arnault, 2009; Cauce et al., 2002; Goldsmith, Jackson, & Hough, 1988; Rickwood, Deane, Wilson, & Ciarrochi, 2005; Shaw et al., 2008). Indeed, research has shown that symptom duration, the number of symptoms experienced and judgements of symptom severity are all significant predictors of health help-seeking (e.g., Elliott, McAteer & Hannaford, 2011; Scott & Walter, 2010).

It is argued here that people may make inaccurate appraisals about their symptoms when using social comparison to make these appraisals, leading to inaccurate judgements about whether or not they are ill and subsequent inaccurate help-seeking behaviour. It is therefore hypothesised that, when appraising their symptoms, people will compare their experience of the symptoms to other people's experience of the same symptoms (Suls, Martin, & Leventhal, 1997). This may lead to inaccurate appraisals and subsequent help-seeking when the distribution of the comparison dimension (e.g., symptom frequency) within the comparison sample is not representative of the distribution of the dimension in the general population. Specifically, it is proposed that people will compare the absolute experience of their symptoms, such as their frequency and duration, to the frequency/duration with which they think other people have experienced the symptoms. It is hypothesised that this comparison will directly influence help-seeking behaviours as absolute aspects of symptoms such as their duration have been shown to predict help-seeking behaviour directly (e.g., Elliott et al., 2011). Where the symptom being appraised is indicative of a serious underlying disorder, the comparison may influence judgements about whether or not this disorder is present (as seen in Melrose et al., 2013) and these judgements may also influence help-seeking behaviour. The comparison may also indirectly influence help-seeking when it is used to make judgements known to influence help-seeking such as judgements of symptom severity (e.g., Elliott et al., 2011). Help-seeking may also be indirectly influenced by the comparison if the comparison evokes distress as

there is a strong link between affective responses to symptom appraisals and help-seeking (Martin, Rothrock, Leventhal, & Leventhal, 2003). Figure 3.1 below outlines these associations, which are tested in the two studies outlined in this Chapter.

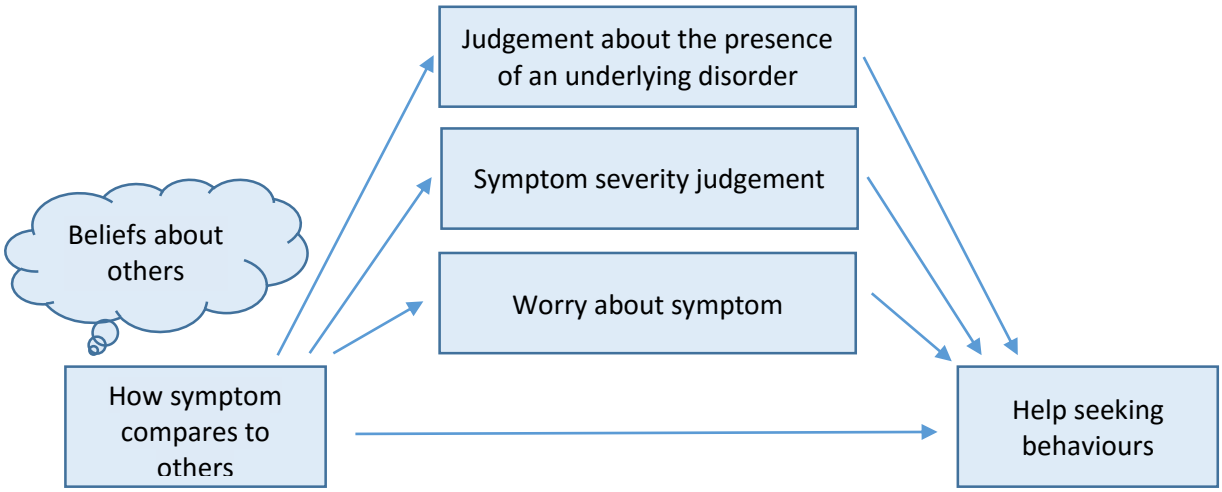

Figure 3.1. Hypothesised direct and indirect associations between social comparison and help-seeking.

Social Comparison and Help-seeking

To my knowledge, these specific associations have not been tested before in the social comparison literature. Previous research on the use of social comparison when making decisions about seeking help for health problems has focused on the use of the 'lay referral structure' (Freidson, 1961) for information and advice about health help-seeking. When people experience ambiguous somatic or psychological sensations they use social comparison to reduce uncertainty and to appraise the significance of what they are feeling (Safer et al., 1979; Suls, 2011). They speak to people that are similar to them in age, gender, risk behaviours and physical shape (Buunk, Gibbons, & Visser, 2002) – the 'lay referral structure' – in order to better understand their symptoms and to get advice about what to do about them. It is proposed that incorrect advice from others may cause people to seek help when they do not need it or not seek help when they do need it (Buunk et al., 2002). This effect of the lay referral structure on help-seeking was highlighted in a classic study where participants were asked to imagine that they were experiencing various

symptoms and to rate, on a nine-point scale from 1 = “very unlikely” to 9 = “very likely”, how likely they would be to visit a doctor about the symptoms (Sanders, 1981). Participants were also given an objective recommendation ascertained from a fictitious physical test as to whether or not they should see a doctor along with social comparison information that either contrasted or supported the objective test recommendation. As would be expected, when both pieces of information made the same recommendation, participants’ decisions to seek help were congruent with the recommendations (average likelihood rating following a recommendation to seek help: 8.5, and not seek help: 1.6). Participants whose test results indicated that they should see a doctor were unaffected by the recommendations of friends and family in the scenario not to seek help (average likelihood rating: 8.5). However, when the test indicated that they should not seek help but friends and family thought they should, participants still decided that they would likely seek help (average likelihood rating: 8.4). This demonstrates how social comparison information[3] can sometimes override objective information and lead to inaccurate decisions.

This previous work on the effect of social comparison on help-seeking is slightly different to the investigation being carried out here. Although Suls et al. (1997) acknowledge that people may compare their experience of symptoms to that of others in order to make help-seeking decisions, little evidence of the testing of this proposition could be found. One study that does provide some empirical evidence for this hypothesis found, using data from a large US national survey (*N* = 36,679), that people who stated that they were more worried, nervous, or anxious than others were significantly more likely to have sought help for mental health problems in the last 12 months (operationalised as whether or not they saw a doctor or mental health professional for problems with emotions, nerves

[3] The “social comparison” information given was whether or not friends or family recommended a visit to the doctor. In my opinion this is not social comparison as no comparison took place - the symptoms were not compared to other people’s symptoms, the friends and family simply gave their opinion. However, this is classed in both the paper and the social comparison literature as social comparison which is why it is referred to as such here.

or mental health, were hospitalised overnight or prescribed medication for these problems) than those who believed they were less worried, nervous, or anxious than others (Mojtabai, 2008). Overall, 31.7% of people who stated that they were more worried, nervous, or anxious than others sought help whereas only 8.4% of people who said that they were less worried, nervous or anxious than others did. Furthermore, this effect of compared distress held when socio-demographic and clinical variables, including actual mental distress (measured using the K10 scale: Furukawa, Kessler, Slade, & Andrews, 2003, which is designed to measure non-specific psychological distress in the general population), that were significantly associated with both compared distress and help-seeking, were controlled in the analysis. The author argues that the study provides preliminary evidence that people's perceptions about how their mental distress compares to others influenced their mental health help-seeking behaviour and that variations in the evaluation of experienced distress may contribute to inaccurate help-seeking behaviour. Although not explicitly mentioned in the paper, some evidence of this is shown in the study. Mojtabai (2008) shows a graph of the percentage of participants that sought help across the two compared distress groups (those that rated themselves as more worried, nervous, or anxious than others and those who rated themselves as less so) stratified across levels of actual distress – four score ranges on the K10 (<10, 10-19, 20-29, 30-40). Differences in help-seeking between the two compared distress groups were significant across all levels of actual distress – people who thought they were more worried, nervous, or anxious than others were more likely to seek help than those who thought they were less worried, nervous, or anxious than others. Kessler et al. (2002) state that people scoring under 20 on the K10 are likely to be well. This means that, of the people in the study likely to not have a mental health problem but who sought help anyway, significantly more thought they were more worried, nervous, or anxious than others. This indicates (although obviously no direct conclusions can be drawn) that it is possible that there is a link between inaccurate social

comparisons and inaccurate help-seeking behaviour. Unfortunately it is not known whether those people in the study likely to have a mental health problem but who did not seek help were more likely to think they were less worried, nervous, or anxious than others. The studies outlined below investigate differences in beliefs about how one compares to others both in people who are likely to have sought help when it was not needed and in people who have not sought help when it was needed.

Current Studies

The main aims of the current studies, outlined below, are therefore to:

1. Replicate the findings of Melrose et al. (2013), namely the association between how symptoms of depression and anxiety are believed to compare to others and perceptions of the presence of these disorders (Study 1),
2. Test for associations between beliefs about how symptoms compare to others and other constructs related to health help-seeking, namely judgements of symptom severity and worry about symptoms (Study 1 and 2),
3. Test for direct and indirect associations between beliefs about how symptoms compare to others and help-seeking as outlined in Figure 3.1 (Study 1 and 2),
4. Investigate how people compare to others when making health-related judgements and decisions about help-seeking, namely whether they use rank or average-based comparisons as consistent with DbS and ALT respectively (Study 1 and 2), and
5. Explore differences in beliefs about others and the potential impact these differences may have on the accuracy of help-seeking decisions (Study 1).

Study 1: Mental Health

In the original Melrose et al. (2013) study, participants were asked about the number of days a month they felt depressed and the number of days a month they felt anxious (depression and anxiety symptom occurrence). They were then given two distribution

elicitation tasks (explained in Chapter 2) asking them to estimate on how many days a month different percentiles of the general population feel depressed and feel anxious (from those who experience these feelings the most to those who experience them the least). From these answers, the relative ranked position of each participant's symptom occurrence within their elicited distribution of symptom occurrence was calculated along with the distance of their occurrence from the mean of the distribution (i.e., the difference in days between their occurrence and the mean occurrence in their elicited distribution). These values were then used to predict participant's answers to a question asking whether they thought they had depression and/or an anxiety disorder (on a 5-point scale from "definitely not" to "definitely"). The results (reported in Table 3.4 of the results section below for comparison to the current study) showed that participants' judgements of whether they thought they suffered from depression or an anxiety disorder were predicted by the ranked position of their depression and anxiety symptom occurrence within their believed distribution of symptom occurrence in the general population. This was over and above their actual symptom occurrence (the number of days a month they experienced the symptoms) and the distance of this occurrence from the mean of their distributions.

As previously stated, the main aims of Study 1 were to replicate the above findings and extend them by investigating whether there are any associations between beliefs about how symptoms compare to others, judgements of symptom severity and worry about symptoms and help-seeking behaviours. As is outlined in the method section below, a very similar methodology to the original study was used. Participants were again asked about the frequency with which they experienced feelings of depression and anxiety and a distribution elicitation task was used to calculate their rank within and distance from the mean of their believed distribution of symptom occurrence within the general population. In addition to this, the new rank and average measures outlined in Chapter 2 were also used allowing for a comparison of the two methodologies. Participants were again asked

about whether they thought they had depression and/or an anxiety disorder and they were also asked how severe they thought their symptom occurrence was and how worried they were about it. The current study builds on the original by investigating whether social comparisons influenced participants' help-seeking behaviour and perceptions about their symptoms. Diagnostic measures were also included to investigate the accuracy of help-seeking decisions and any potential effects of social comparison on this decision.

Method

Participants

The 643 participants that completed this study were recruited through CrowdFlower panels and had a mean age of 35.0 years (*SD* = 11.7, range: 18-94 years), were predominantly White (85%, Indian = 3%, Black = 2%, Chinese = 2%, Pakistani = 2%, Other = 6%) and 48% were male. Participants were all resident in the U.K. (as in the original study) and took part from towns and cities all over the country. Participants were mainly educated up to university (44%) or post-secondary (i.e., education between 16 and 18 years of age such as A-levels: 34%) level (some high school = 5%, finished school at 16 = 17%, rather not say < 1%). Participants were mainly low earners (< £14,999 = 33%, £15,000-£24,999 = 26%, £25,000-£34,999 = 16%, £35,000-£44,999 = 8%, >£45,000 = 7%, rather not say = 10%) and received $0.75 on completion of the study which took 10-15 minutes to complete.

Design and procedure

The questionnaire that participants completed in this study was designed and hosted using Qualtrics. Participants were asked the following:

Symptom occurrence: Participants' experience of feeling depressed and anxious was obtained through asking: "in a typical month, on how many days do you feel down, depressed or hopeless?" and "in a typical month, on how many days do you feel nervous, anxious or on edge?". This wording of symptoms was taken from two of the four PHQ-4 (Kroenke, Spitzer, Williams, & Löwe, 2009) items. The PHQ-4 is a brief screening tool for

anxiety and depression. In all questions where participants were asked to give an answer of "the number of days a month", they were asked to consider a month to be 30 days.

Distribution elicitation task: Participants were asked to imagine that all of the adults in the UK were lined up in order of the number of days a month they experience a symptom from those that experience it the least to those that experience it the most. They were then asked how many days a month they thought that adults at various positions along this line (corresponding to the 5th, 20th, 50th, 80th, and 95th percentiles of the distribution) experience the symptom. They were asked to do this for the two symptoms outlined above: "feeling down, depressed or hopeless" and "feeling anxious, nervous or on edge". Figure 3.2 below shows how this task was presented to participants.

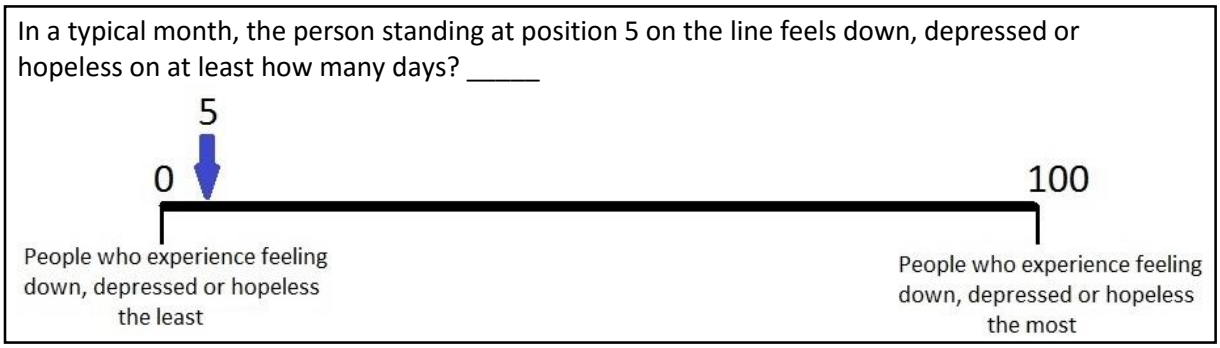

Figure 3.2. An example of one of the distribution elicitation task questions.

Direct measure of rank of symptom occurrence within the general population: After answering the above distribution elicitation questions participants were again shown the line that went from "people experiencing [symptom] the least" to "people experiencing [symptom] the most" and were asked what position on the line they themselves occupied for each symptom separately. This gave a direct measure of where each participant believed their symptom occurrence ranked in comparison to others.

Direct measure of the average symptom occurrence in the general population: Participants were asked what they thought the average occurrence of each symptom was in the following format: "thinking about the general adult population, what do you think the average number of days that people feel down, depressed or hopeless is in a typical

month?" and "thinking about the general adult population, what do you think the average number of days that people feel nervous, anxious or on edge is in a typical month?".

The following outcomes were measured:

Whether or not they thought they had depression and/or anxiety: These questions were the same as in the original study: "do you think you suffer from depression?" and "do you think you suffer from anxiety?" each answered on a 5-point scale: "definitely not", "probably not", "maybe", "probably", and "definitely".

Worry: Participants rated how worried they were about the amount of time they experience each symptom, e.g.: "how worried are you about the amount of time you feel down, depressed or hopeless?" on a slider scale from "not at all" to "very much".

Symptom severity: Participants then answered the questions "how severe are your feelings of depression?" and "how severe are your feelings of anxiety?" on a slider scale from "not at all severe" to "extremely severe".

For all questions where a slider scale was used Qualtrics records where the participant placed the slider as a number between 0 and 100.

Help-seeking: For each symptom, participants were asked: "which of these actions have you taken over the past year to manage [symptom], tick all that apply" and were provided with the following list: looked for information, discussed feelings with other people, phoned NHS 24/NHS Direct/NHS 111, consulted a mental health charity or support group, consulted a pharmacist, consulted a nurse or doctor, consulted a therapist, took complementary medication, took conventional medication, or none of the above. These behaviours were taken and adapted from Elliott et al. (2011) who carried out a U.K. wide study on how people respond to symptoms.

Depression screening: Participants completed the Patient Health Questionnaire-9 (PHQ-9: Kroenke & Spitzer, 2002, see Appendix 1) which is both a diagnostic tool and a measure of symptom severity for depression and has been validated for use within the

general population (Martin, Rief, Klaiberg, & Braehler, 2006). Items are based on diagnostic criteria and respondents rate the frequency with which they have been bothered by nine depression-related problems over the last 2 weeks on a 4-point scale: “not at all”, “several days”, “more than half the days”, and “nearly every day”. Items are scored between 0 and 3 with total scores between 0 and 27. A total score of 1-4 = no, 5-9 = mild, 10-14 = moderate, 5-19 = moderately severe and 20-17 = severe depression. The authors recommend that a cut point of 10 be used for screening as this has a sensitivity for major depression of 88% and a specificity of 88%. The PHQ-9 has excellent internal (Cronbach's α = 0.89) and test-retest (*r* = .84) reliability (Kroenke, Spitzer, & Williams, 2001). Scores on the PHQ-9 have been shown to correlate highly and significantly with both Beck Depression Inventory scores (*r* = .71; the BDI is another well used measure of depression severity) and scores on the mental health subscale of the SF-36 (*r* = -.71, lower scores on this subscale indicate poorer mental health, hence the negative correlation) suggesting that the PHQ-9 has good convergent validity also (Martin et al., 2006).

Anxiety screening: Participants completed the Generalised Anxiety Disorder-7 (GAD-7: Spitzer, Kroenke, Williams, & Löwe, 2006, see Appendix 2) which is both a diagnostic tool and a measure of symptom severity and has been validated for use within the general population (Löwe et al., 2008). Although the GAD-7 measures generalised anxiety disorder symptoms, it is a good overall measure of anxiety and is sensitive to posttraumatic stress disorder, panic disorder, and social anxiety disorder, which are the three most prevalent anxiety disorders seen in patients with somatic diseases (Kroenke, Spitzer, Williams, Monahan, & Löwe, 2007). Just as with the PHQ-9, items are based on diagnostic criteria and respondents rate the frequency with which they have been bothered by seven anxiety-related problems using the same scale as the PHQ-9. Total scores range between 0 and 21 with 0-4 = minimal, 5-9 = mild, 10-14 = moderate and 15-21 = severe anxiety. As with the PHQ-9, the authors recommend that a cut point of 10 be used for screening as this has a

sensitivity for anxiety of 89% and a specificity of 82%. The GAD-7 also has excellent internal (Cronbach's α = 0.92) and test-retest (r = .83) reliability (Spitzer et al., 2006). GAD-7 scores correlate highly and significantly with the Beck Anxiety Inventory (r = .72) and the anxiety subscale of the Symptom Checklist-90 (r = .74) indicating good convergent validity (Spitzer et al., 2006).

Questions were ordered into three blocks: questions on symptom occurrence, comparison questions (distribution elicitation questions and direct rank and average questions) and outcome measures. The order that these blocks were presented was counterbalanced so that half the participants saw the outcome measures first and half saw them last and then half saw the symptom occurrence questions before the comparison questions and half saw the comparison questions before the symptom occurrence questions. Demographic characteristics (age, gender, level of education, ethnicity, and income bracket) were collected right at the end of the survey after all the above questions were presented.

Results

Data for each symptom were analysed separately. This results section addresses each of the aims outlined in the introduction on page 54 in turn although the use of rank or average-based comparison strategies (aim 4) is discussed throughout. This is achieved, in the main, through regression analyses and the main independent variables are labelled as follows:

Symptom occurrence: There was one measure of symptom occurrence for each symptom: the number of days in the last month the participant felt down, depressed or hopeless (*M* = 8.92, *SD* = 8.52) and the number of days in the last month they felt nervous, anxious or on edge (*M* = 9.69, *SD* = 9.02).

Rank: Two rank measures were calculated from the questions outlined above. Answers to the distribution elicitation task were used to construct each participant's beliefs

about the distribution of depression and anxiety symptom occurrence within the general population (their believed distribution). Using their answers to the symptom occurrence questions, the rank of each participant's symptom occurrence within their believed distribution of symptom occurrence was calculated, where possible, for both depression and anxiety (see Chapter 2 for a detailed explanation of how this is carried out). Due to inevitable exclusions from participants not understanding the task (again see Chapter 2 for more detail) this resulted in subjective rank values between 0 and 1 for 549 participants included in the depression analyses (M = 0.31, SD = 0.31) and 563 participants included in the anxiety analyses (M = 0.32, SD = 0.32). This variable is labelled 'indirect rank' in the analyses outlined below. The variable 'direct rank' is the direct measure of where participants think their symptom occurrence ranks in comparison to others, measured by the position they selected on the line that went from "people who experience the symptom the least" to "people who experience the symptom the most". The resulting direct rank values are between 0 and 100 (depression: M = 32.43, SD = 31.51, anxiety: M = 33.40, SD = 31.88). For both rank measures, high scores indicate that the participant believes their occurrence ranks highly within the general population, i.e., that the majority of people experience symptoms less frequently than they do.

Distance from the average: Again, there are two distance from the average variables calculated from the two different methodologies. The mean of each participant's believed distribution of depression and anxiety symptom occurrence in the general population was calculated and deducted from their own symptom occurrence to create the 'indirect distance from the average' variable (depression: M = -4.97, SD = 8.45, anxiety: M = -4.52, SD = 8.92). Participants' answers to the two questions asking what they though the average occurrence of each symptom in the general population was were deducted from their answers to the relevant symptom occurrence questions. This gave a direct measure of how much their experience of the symptoms differed from their believed average experience of

others. This variable is labelled 'direct distance from the average' (depression: *M* = 0.26, *SD* = 8.84, anxiety: *M* = -0.03, *SD* = 9.46). Both distance from the average variables have possible values ranging from -30 to 30 with negative values indicating that the participant experiences the symptom less often than what they believe the average person does and positive values indicating that the participant experiences the symptom more often.

Analyses: As explained in the methodology chapter on pages 41-45, when regression was used to investigate the aims of the study, models with the following steps were formulated:

- Step 1: includes age, gender and symptom occurrence,
- Step 2a: includes age, gender, symptom occurrence and rank,
- Step 2b: includes age, gender, symptom occurrence and the distance from the average[4].

This is because it is hypothesised that people will either use rank-based comparison processes or will compare to the average of the comparison group when making judgements[5]. Evidence for the use of either rank or average-based social comparison is obtained from the variable's performance in the model and from comparing the fit of the step 2a and 2b models (non-nested models). This is done using Hotelling's t-tests (multiple linear regression models) and by calculating Bayesian information criterion (BIC) for the models (binary and ordinal regression models).

Aim 1

The first aim of the study was to see whether the findings from Melrose et al. (2013) could be replicated using rank and distance from the average variables calculated from both methodologies – the standard distribution elicitation task as used in the original study

[4] This format was used in all regression analyses reported in this thesis apart from those reported under Aim 1 of the current study for reasons explained in the paragraph below.
[5] Models which included all five independent variables (age, gender, absolute experience, rank and distance from the average) were also formulated. The pattern of results seen did not differ between methods.

and the new, simpler, direct comparison questions. Unfortunately, as Table 3.1 shows, both the indirect rank and indirect distance from the average variables were highly correlated with symptom occurrence for both symptoms. This meant that collinearity would likely be an issue in the regression analyses when these variables were included together in the models. Collinearity statistics confirmed this to be the case - tolerance values for these variables obtained through running multiple linear regression analyses ranged from .075 to .107. Although the correlations between symptom occurrence and the direct rank and distance from the average variables for both symptoms were also high (but not as high as the indirect measures), the tolerance values for these variables were acceptable, ranging from .351 to .416. It was therefore decided that the indirect measures would be used to investigate comparison effects for the first study aim only as this measure had been used in the original study. The analyses using these variables have been adjusted accordingly (see Table 3.2); the step 2a and 2b models do not include symptom occurrence and because they are no longer additional steps to the original step 1 model the three models calculated are labelled models 1-3. For investigation of the rest of the study aims, only the direct measures are used in analyses.

Tables 3.2 and 3.3 below show both the ordinal and multiple linear regression coefficients for predictors of participants' judgements about whether they thought they had depression or anxiety using the indirect and direct measures respectively. To recap, this outcome was measured on a 5-point scale from 1 = "definitely not" to 5 = "definitely" as in the original study (depression: *M* = 2.85, *SD* = 1.37, anxiety: *M* = 3.06, *SD* = 1.34). There was a good range of responses despite the non-clinical sample: depression: "definitely not" = 19.3%, "probably not" = 26.1%, "maybe" = 21.8%, "probably" = 15.2%, "definitely" = 17.6%; anxiety: "definitely not" = 13.7%, "probably not" = 26.0%, "maybe" =

21.2%, “probably” = 19.0%, “definitely” = 20.2%. For comparison, the results from the original study are outlined in Table 3.4 below.

Table 3.1.
Correlations between Continuous Independent Variables.

	Depression							
	Age	Symptom Occurrence	Indirect Rank	Indirect Distance from the Average	Age	Symptom Occurrence	Direct Rank	Direct Distance from the Average
Age	1	-.091*	-0.077	-0.075	1	-0.076	-.099*	-0.069
Symptom Occurrence	-.091*	1	.958***	.945***	-0.076	1	.804***	.774***
Rank	-0.077	.958***	1	.967***	-.099*	.804***	1	.648***
Distance from the Average	-0.075	.945***	.967***	1	-0.069	.774***	.648***	1
N	549	549	549	549	643	643	643	643
	Anxiety							
Age	1	-.109**	-.098*	-.094*	1	-.087*	-.134***	-0.061
Symptom Occurrence	-.109**	1	.962***	.948***	-.087*	1	.778***	.764***
Rank	-.098*	.962***	1	.968***	-.134***	.778***	1	.624***
Distance from the Average	-.094*	.948***	.968***	1	-0.061	.764***	.624***	1
N	563	563	563	563	643	643	643	643

Note.*$p < .05$, **$p < .01$, ***$p < .001$ (two-tailed).

Table 3.2.

Ordinal and Multiple Linear Regression Analyses Predicting Whether Participants think they have Depression and Anxiety from Age, Gender, Symptom Occurrence, Indirect Rank of Symptom Occurrence, and Indirect Distance from the Believed Average Symptom Occurrence of People in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

		Ordinal Regression				Multiple Linear Regression				
Predictors: Depression	R_N^2	Estimate (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	R^2	B (95% CI)	Std. Error	β	*p*
Model 1	.52			**<.001**		.52***				**<.001**
Age		0.00 (-0.02-0.01)	0.01	.670	1.00 (0.98-1.01)		0.00 (-0.01-0.00)	0.00	-.02	.519
Gender		**0.54 (0.22-0.85)**	**0.16**	**.001**	**1.71 (1.24-2.35)**		**-0.31 (-0.48--0.14)**	**0.08**	**-.11**	**.001**
Symptom Occurrence		**0.22 (0.19-0.25)**	**0.01**	**.001**	**1.25 (1.21-1.28)**		**0.11 (0.10-0.12)**	**0.01**	**.70**	**.001**
Model 2	.47			**<.001**		.47***				**<.001**
Age		-0.01 (-0.02-0.01)	0.01	.380	0.99 (0.98-1.01)		0.00 (-0.01-0.00)	0.00	-.03	.319
Gender		**0.54 (0.23-0.86)**	**0.16**	**.001**	**1.72 (1.26-2.36)**		**-0.33 (-0.49--0.18)**	**0.08**	**-.12**	**.001**
Indirect Rank		**5.31 (4.63-5.98)**	**0.35**	**.001**	**201.58 (102.39-396.84)**		**2.96 (2.66-3.24)**	**0.14**	**.67**	**.001**
Model 3	.47			**<.001**		.46***				**<.001**
Age		-0.01 (-0.02-0.01)	0.01	.382	0.99 (0.98-1.01)		0.00 (-0.01-0.00)	0.00	-.04	.270
Gender		**0.57 (0.25-0.88)**	**0.16**	**.001**	**1.76 (1.28-2.41)**		**-0.34 (-0.52--0.17)**	**0.09**	**-.12**	**.001**
Indirect Distance from the Average		**0.19 (0.17-0.22)**	**0.01**	**.001**	**1.21 (1.18-1.24)**		**0.11 (0.10-0.12)**	**0.01**	**.66**	**.001**
Predictors: Anxiety										
Model 1	.54			**<.001**		.51***				**<.001**
Age		0.00 (-0.02-0.01)	0.01	.644	1.00 (0.98-1.01)		0.00 (-0.01-0.00)	0.00	-.02	.537
Gender		**0.40 (0.08-0.71)**	**0.16**	**.014**	**1.49 (1.08-2.04)**		**-0.20 (-0.38--0.02)**	**0.08**	**-.07**	**.015**
Symptom Occurrence		**0.22 (0.20-0.25)**	**0.01**	**.001**	**1.25 (1.22-1.28)**		**0.11 (0.10-0.11)**	**0.00**	**.70**	**.001**
Model 2	.48			**<.001**		.47***				**<.001**
Age		-0.01 (-0.02-0.01)	0.01	.376	0.99 (0.98-1.01)		0.00 (-0.01-0.00)	0.00	-.03	.318
Gender		**0.38 (0.07-0.70)**	**0.16**	**.017**	**1.47 (1.07-2.01)**		**-0.21 (-0.39--0.04)**	**0.09**	**-.08**	**.011**
Indirect Rank		**5.40 (4.71-6.09)**	**0.35**	**.001**	**221.29 (111.51-439.16)**		**2.80 (2.56-3.03)**	**0.12**	**.67**	**.001**
Model 3	.49			**<.001**		.47***				**<.001**
Age		-0.01 (-0.02-0.01)	0.01	.296	0.99 (0.98-1.01)		0.00 (-0.01-0.00)	0.00	-.03	.257
Gender		**0.43 (0.11-0.74)**	**0.16**	**.008**	**1.53 (1.12-2.10)**		**-0.24 (-0.41--0.07)**	**0.08**	**-.09**	**.008**
Indirect Distance from the Average		**0.20 (0.17-0.22)**	**0.01**	**.001**	**1.22 (1.19-1.25)**		**0.10 (0.09-0.11)**	**0.00**	**.67**	**.001**

Note. n = 549 for depression and n = 563 for anxiety, CI = confidence interval, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. ***p <. 001.

Table 3.3.

Ordinal and Multiple Linear Regression Analyses Predicting Whether Participants think they have Depression and Anxiety from Age, Gender, Symptom Occurrence, and Direct Social Comparison Measures (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

		Ordinal Regression				Multiple Linear Regression				
Predictors: Depression	ΔR_N^2	Estimate (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.49			**<.001**		.48***				**<.001**
Age		0.00 (-0.02-0.01)	0.01	.563	1.00 (0.98-1.01)		0.00 (-0.01-0.00)	0.00	-.02	.463
Gender		**0.39 (0.09-0.68)**	**0.15**	**.010**	**1.47 (1.10-1.97)**		**-0.23 (-0.39--0.08)**	**0.08**	**-.08**	**.002**
Symptom Occurrence		**0.21 (0.18-0.23)**	**0.01**	**.001**	**1.23 (1.20-1.26)**		**0.11 (0.10-0.12)**	**0.01**	**.68**	**.001**
Step 2	.08			**<.001**		.09***				**<.001**
Age		0.00 (-0.01-0.01)	0.01	.930	1.00 (0.99-1.01)		0.00 (-0.01-0.01)	0.00	-.00	.931
Gender		**0.44 (0.14-0.73)**	**0.15**	**.004**	**1.55 (1.15-2.08)**		**-0.21 (-0.36--0.09)**	**0.07**	**-.08**	**.001**
Symptom Occurrence		**0.10 (0.07-0.13)**	**0.02**	**.001**	**1.10 (1.07-1.14)**		**0.05 (0.03-0.06)**	**0.01**	**.29**	**.001**
Direct Rank		**0.04 (0.04-0.05)**	**0.00**	**.001**	**1.04 (1.04-1.05)**		**0.02 (0.02-0.03)**	**0.00**	**.49**	**.001**
Step 2	.01			**<.001**		.00*				**<.001**
Age		0.00 (-0.02-0.01)	0.01	.596	1.00 (0.98-1.01)		0.00 (-0.01-0.00)	0.00	-.02	.470
Gender		**0.41 (0.11-0.70)**	**0.15**	**.007**	**1.50 (1.12-2.02)**		**-0.24 (-0.39--0.09)**	**0.08**	**-.09**	**.001**
Symptom Occurrence		**0.18 (0.15-0.21)**	**0.02**	**.001**	**1.19 (1.16-1.23)**		**0.10 (0.08-0.11)**	**0.01**	**.60**	**.001**
Direct Distance from the Average		**0.04 (0.02-0.07)**	**0.01**	**.002**	**1.04 (1.02-1.07)**		**0.02 (0.00-0.03)**	**0.01**	**.11**	**.032**
Predictors: Anxiety										
Step 1	.51			**<.001**		.48***				**<.001**
Age		-0.01 (-0.02-0.01)	0.01	.308	0.99 (0.98-1.01)		0.00 (-0.01-0.00)	0.00	-.04	.229
Gender		**0.34 (0.04-0.63)**	**0.15**	**.026**	**1.40 (1.04-1.88)**		**-0.17 (-0.33--0.02)**	**0.08**	**-.07**	**.027**
Symptom Occurrence		**0.21 (0.18-0.23)**	**0.01**	**.001**	**1.23 (1.20-1.26)**		**0.10 (0.09-0.11)**	**0.00**	**.68**	**.001**
Step 2	.05			**<.001**		.06***				**<.001**
Age		0.00 (-0.01-0.01)	0.01	.865	1.00 (0.99-1.01)		0.00 (-0.01-0.00)	0.00	-.01	.723
Gender		**0.44 (0.15-0.74)**	**0.15**	**.004**	**1.56 (1.16-2.10)**		**-0.20 (-0.36--0.05)**	**0.08**	**-.07**	**.014**
Symptom Occurrence		**0.13 (0.10-0.16)**	**0.01**	**.001**	**1.14 (1.10-1.17)**		**0.06 (0.04-0.07)**	**0.01**	**.37**	**.001**
Direct Rank		**0.03 (0.03-0.04)**	**0.00**	**.001**	**1.03 (1.03-1.04)**		**0.02 (0.01-0.02)**	**0.00**	**.40**	**.001**
Step 2	.00			**<.001**		.01**				**<.001**
Age		-0.01 (-0.02-0.01)	0.01	.277	0.99 (0.98-1.01)		0.00 (-0.01-0.00)	0.00	-.04	.195
Gender		**0.34 (0.04-0.64)**	**0.15**	**.025**	**1.40 (1.04-1.89)**		**-0.17 (-0.34--0.01)**	**0.08**	**-.06**	**.033**
Symptom Occurrence		**0.18 (0.15-0.21)**	**0.02**	**.001**	**1.19 (1.16-1.23)**		**0.09 (0.08-0.10)**	**0.01**	**.59**	**.001**
Direct Distance from the Average		**0.04 (0.02-0.07)**	**0.01**	**.001**	**1.04 (1.02-1.07)**		**0.02 (0.00-0.03)**	**0.01**	**.12**	**.007**

Note. N = 643, CI = confidence interval, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. **p* < .05, ***p* < .01, ****p* <. 001.

Table 3.4.
Regression Coefficients from the Original Study.

	Think they suffer from depression				Think they suffer from anxiety			
Predictor Variable	Estimate	Std. Error	Wald	*p*	Estimate	Std. Error	Wald	*p*
Age	-0.019	0.013	2.225	.136	-0.003	0.013	0.038	.844
Gender	-0.027	0.331	0.007	.934	-0.368	0.345	1.141	.285
Symptom occurrence	**0.176**	**0.028**	**39.117**	**.000**	**0.210**	**0.032**	**42.545**	**.000**
Nagelkerke Pseudo-R^2		**.367**				**.417**		
Age	-0.025	0.013	3.457	.063	-0.006	0.013	0.231	.631
Gender	0.162	0.336	0.231	.630	-0.289	0.345	0.700	.403
Symptom occurrence	0.064	0.041	2.515	.113	**0.106**	**0.044**	**5.657**	**.017**
Subjective rank	**3.303**	**0.924**	**12.764**	**.000**	**3.070**	**0.996**	**9.505**	**.002**
Nagelkerke Pseudo-R^2		**.429**				**.460**		
Age	-0.020	0.013	2.287	.130	-0.006	0.014	0.210	.647
Gender	0.023	0.333	0.005	.945	-0.283	0.350	0.655	.418
Symptom occurrence	0.194	0.031	39.350	.000	0.225	0.035	41.530	.000
Subjective mean	-0.064	0.042	2.293	.130	-0.057	0.044	1.694	.193
Nagelkerke Pseudo-R^2		**.379**				**.425**		

Note. n = 133 for depression, n = 135 for anxiety. Significant models and predictors are highlighted in bold. Taken from Melrose et al. (2013, p178).

As reported in Chapter 2, the two different rank and distance from the average measures correlated well together (depression rank variables: r = .843, anxiety rank variables: r = .814, depression distance from the average variables: r = .748, and anxiety distance from the average variables: r = .748, all p values < .001) indicating good construct validity. Reassuringly, although the analysis is slightly different, Tables 3.2 and 3.3 show a very similar pattern of results providing further evidence that the different rank and distance from the average variables are measuring the same constructs. Therefore, only the findings from Table 3.3 are discussed, as this includes the full analysis with participants' absolute experience of the symptoms in the step 2a and 2b models. This is more reflective of how we assume people actually make judgements and the models reported in Table 3.3 are directly comparable to the findings from the original study reported in Table 3.4.

Comparison of the step 1 models in Tables 3.3 and 3.4 shows that, in both studies, the number of days in a typical month that participants experience feeling depressed or anxious is a significant predictor of whether they think they have depression or anxiety

respectively. In the original study, this effect was removed when the rank of this symptom occurrence within the general population was added to the model in step 2a for depression and was attenuated for anxiety. In the current study, the effect of symptom occurrence is not removed when rank is added to the model but it is attenuated for both symptoms by around 50%. The ordinal regression coefficients for symptom occurrence reduce from 0.21 (depression and anxiety step 1) to 0.10 (depression step 2a) and 0.13 (anxiety step 2a) when rank is added to the model. Rank becomes the strongest predictor of whether participants think they are depressed - comparison of the beta coefficients in the step 2a multiple linear regression models showed that the rank beta weight was significantly larger than the symptom occurrence beta weight. However, symptom occurrence and rank predict whether participants think they have anxiety equally well - there was no significant difference in the beta coefficients for these variables. In the original study, the distance of participant's symptom occurrence from the mean of their believed distribution of symptom occurrence in the population was not a significant predictor of whether they thought they had depression or anxiety. This is not the case for the current study where it was a significant predictor in both cases, as can be seen from the step 2b models in Table 3.3. However, for both symptoms, distance from the average was a significantly weaker predictor than both symptom occurrence and rank. This can be seen from comparison of the beta coefficients for these variables and the amount of additional variance accounted for by the step 2a and 2b models. The addition of rank to the step 1 models resulted in an 8% increase (significant) in explained variance in judgements of whether participants thought they were depressed and a 5% increase (significant) in explained variance in judgements of whether participants thought they were anxious. Addition of the distance from the average variables to the step 1 models resulted in less than a 1% increase (only significant in the anxiety model) in explained variance in both judgements. Comparison of the ordinal regression models showed that, for both symptoms, there was very strong

evidence that the step 2a models including age, gender, symptom occurrence and rank fitted the data much better than the step 2b models including age, gender, symptom occurrence and distance from the average (differences in -2LL between the step 2a and step 2b models: depression models = -95.03, anxiety models = -61.64). Although not of primary interest, it should also be stated that there is a significant effect of gender such that females were significantly more likely to say that they thought they had depression or anxiety than men in the current study that was not present in the original.

Summary: Overall, both the original and replication study provide evidence that people compare their experience of depression and anxiety symptoms to other people's experience of these symptoms when making judgements about whether or not they have depression or anxiety respectively. How participants' believed their symptom occurrence compared to that of other people in the general population significantly predicted these judgements. The evidence across studies suggests that when comparing to others, people use rank-based strategies rather than comparing to the average of the comparison sample. Although in the current study both comparison variables were significant predictors of whether participants though they had depression or anxiety, it was the variable measuring where participants believed their symptom occurrence ranked in comparison to others that had the strongest effect. The model comparison showed that there was strong evidence that the models including the rank variables fitted the data better than those including the distance from the average variables. The amount of explained variance that the rank variables accounted for (depression: 9%; anxiety: 6%) was much greater than that of the distance from the average variables (depression: 0%; anxiety: 1%).

Aim 2

The second aim of the study was to test for associations between beliefs about how symptoms compare to others and other constructs related to health help-seeking, namely judgements of symptom severity and worry about symptoms. To recap, these judgements

were both measured on a 0 to 100 scale (depression severity: *M* = 30.95, *SD* = 29.85; anxiety severity: *M* = 33.21, *SD* = 31.55; depression worry: *M* = 36.62, *SD* = 32.97; anxiety worry: *M* = 36.55, *SD* =32.99) with low values indicating mild severity and worry and high values indicating extreme severity and worry. Therefore, multiple linear regression was used to investigate whether the direct rank and distance from the average variables predicted these judgements and these analyses are reported in Tables 3.5 (symptom severity) and 3.6 (worry) below.

Symptom severity: The step 1 models in Table 3.5 show that the number of days a month that participants felt depressed or anxious were significant predictors of how severe they thought their feelings of depression and anxiety were respectively. Age was also a significant predictor in the models but symptom occurrence was by far the strongest predictor of symptom severity judgements (the symptom occurrence beta weights in both models were .72 compared to -.09 for age). These models accounted for 54% of the variance in symptom severity judgements. Addition of the rank variables in the step 2a depression and anxiety models increased the amount of explained variance significantly (by 13% and 14% respectively). For both depression and anxiety, addition of these variables led to an attenuation of the effect of symptom occurrence by around 70%. Where the participants believed their symptom occurrence ranked in comparison to others became the strongest predictor of symptom severity judgements, over and above symptom occurrence (the rank beta weights were significantly greater than the symptom occurrence beta weights). In contrast, for both depression and anxiety, when distance from the average is added to the original step 1 models in step 2b, the increase in the amount of variance in symptom severity judgements accounted for by the models is less than 1% and is only significant in the anxiety model. How much participants' symptom occurrence differs from what they believe the average symptom occurrence of others to be was a significant predictor of anxiety symptom severity judgements only and not depression.

Table 3.5.

Multiple Linear Regression Analyses Predicting Symptom Severity Ratings from Age, Gender, Symptom Occurrence, Direct Rank of Symptom Occurrence, and Direct Distance from the Believed Average *Symptom Occurrence of People in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).*

	Depression					Anxiety				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.54***				**<.001**	.54***				**<.001**
Constant		17.01 (11.41-22.82)	3.03		.001		16.65 (8.86-23.79)	3.75		.001
Age		**-0.24 (-0.38--0.11)**	**0.07**	**-.09**	**.002**		**-0.25 (-0.43--0.07)**	**0.09**	**-.09**	**.007**
Gender		-0.33 (-3.41-3.09)	1.58	-.01	.835		1.73 (-1.25-5.08)	1.70	.03	.311
Symptom Occurrence		**2.52 (2.30-2.75)**	**0.11**	**.72**	**.001**		**2.53 (2.33-2.74)**	**0.10**	**.72**	**.001**
Step 2a	.13***				**<.001**	.14***				**<.001**
Constant		11.28 (6.70-16.57)	2.44		.001		9.39 (3.70-14.76)	2.78		.001
Age		**-0.18 (-0.30--0.07)**	**0.06**	**-.07**	**.004**		**-0.15 (-0.28--0.02)**	**0.07**	**-.06**	**.031**
Gender		0.12 (-2.46-2.55)	1.33	.00	.946		0.87 (-1.75-3.75)	1.42	.01	.562
Symptom Occurrence		**0.83 (0.47-1.24)**	**0.18**	**.24**	**.001**		**0.91 (0.56-1.29)**	**0.17**	**.26**	**.001**
Rank		**0.57 (0.48-0.65)**	**0.05**	**.60**	**.001**		**0.59 (0.49-0.68)**	**0.05**	**.60**	**.001**
Step 2b	.00				**<.001**	.00*				**<.001**
Constant		18.31 (11.96-24.93)	3.28		.001		19.17 (10.90-27.46)	3.87		.001
Age		**-0.24 (-0.38--0.10)**	**0.07**	**-.09**	**.002**		**-0.25 (-0.41--0.09)**	**0.09**	**-.09**	**.006**
Gender		-0.41 (-4.00-2.69)	1.63	-.01	.794		1.77 (-1.80-5.07)	1.73	.03	.305
Symptom Occurrence		**2.37 (2.07-2.69)**	**0.16**	**.68**	**.001**		**2.28 (1.98-2.60)**	**0.15**	**.65**	**.001**
Distance from the Average		0.19 (-0.15-0.47)	0.16	.06	.239		**0.32 (0.01-0.58)**	**0.15**	**.10**	**.039**

Note. N = 643, CI = confidence interval. Significant models and predictors are highlighted in bold. *p < .05, ***p < .001.

Table 3.6.

Multiple Linear Regression Analyses Predicting Ratings of Worry About Symptom Occurrence from Age, Gender, Symptom Occurrence, Direct Rank of Symptom Occurrence, and Direct Distance from the Believed Average *Symptom Occurrence of People in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).*

	Depression					Anxiety				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.50***				**<.001**	.30***				**<.001**
Constant		26.19 (18.74-34.00)	3.63		.001		36.94 (28.25-45.80)	4.21		.001
Age		**-0.34 (-0.49--0.18)**	**0.08**	**-.12**	**.001**		**-0.44 (-0.63--0.26)**	**0.10**	**-.16**	**.001**
Gender		-2.79 (-6.42-0.58)	1.86	-.04	.146		**-5.16 (-9.13--0.85)**	**2.10**	**-.08**	**.019**
Symptom Occurrence		**2.64 (2.38-2.87)**	**0.12**	**.68**	**.001**		**1.96 (1.64-2.26)**	**0.15**	**.51**	**.001**
Step 2a	.11***				**<.001**	.09***				**<.001**
Constant		20.33 (13.72-27.00)	3.25		.001		31.50 (23.23-39.39)	3.85		.001
Age		**-0.28 (-0.42--0.13)**	**0.08**	**-.10**	**.001**		**-0.38 (-0.55--0.22)**	**0.09**	**-.14**	**.001**
Gender		-2.34 (-5.39-0.70)	1.61	-.04	.157		**-4.73 (-8.66--0.72)**	**1.97**	**-.07**	**.017**
Symptom Occurrence		**0.92 (0.53-1.28)**	**0.20**	**.24**	**.001**		0.36 (-0.11-0.77)	0.23	.09	.116
Rank		**0.58 (0.48-0.69)**	**0.05**	**.56**	**.001**		**0.54 (0.43-0.66)**	**0.06**	**.52**	**.001**
Step 2b	.00*				**<.001**	.00				**<.001**
Constant		28.61 (21.64-35.64)	3.59		.001		39.32 (31.11-47.57)	4.19		.001
Age		**-0.34 (-0.50--0.18)**	**0.08**	**-.12**	**.001**		**-0.44 (-0.62--0.26)**	**0.10**	**-.16**	**.001**
Gender		-2.95 (-6.14-0.53)	1.77	-.05	.088		**-5.32 (-9.70--0.82)**	**2.11**	**-.08**	**.011**
Symptom Occurrence		**2.36 (2.01-2.71)**	**0.17**	**.61**	**.001**		**1.68 (1.24-2.12)**	**0.21**	**.44**	**.001**
Distance from the Average		**0.35 (-0.02-0.69)**	**0.17**	**.09**	**.038**		0.34 (-0.05-0.73)	0.20	.09	.088

Note. N = 643, CI = confidence interval. Significant models and predictors are highlighted in bold. *p < .05, ***p < .001.

The effect of the distance from the average variable (β = .10) was significantly weaker than the effect of symptom occurrence (β = .65) in the step 2b anxiety model. Unsurprisingly, comparison of the predictive ability of the step 2a and 2b models showed that the step 2a models accounted for significantly more variance in symptom severity judgements than the step 2b models, depression: $t(640) = 7.83$, $p < .001$, anxiety: $t(640) = 8.32$, $p < .001$.

Worry: This pattern of results was much the same for judgements regarding how worried participants were about the amount of time they felt depressed or anxious. The step 1 models in Table 3.6 show that symptom occurrence was the strongest predictor of worry ratings and these models accounted for 50% and 30% of the variance in ratings of how worried participants were about their depression symptom occurrence and anxiety symptom occurrence respectively. When the rank variables are entered into the models at step 2a, the effect of symptom occurrence in the depression model is attenuated by about 60% and is completely removed from the anxiety model. Rank becomes the strongest predictor of worry ratings for both symptoms – the rank beta weights are significantly larger than those of the other predictors in the models. The amount of explained variance in depression and anxiety worry ratings increases significantly by 11% and 9% respectively. Addition of the distance from the average variables to the step 1 models has little effect - the variance accounted for by the models increases by less than 1% (significant for the depression model only). Distance from the average is a significant predictor of worry ratings in the depression model only and is a significantly weaker predictor (β = .09) compared to symptom occurrence (β = .61). The step 2a models accounted for significantly more variance in worry ratings than the step 2b models, depression: $t(640) = 6.56$, $p < .001$, anxiety: $t(640) = 6.32$, $p < .001$.

Summary: Overall, the results provide evidence that people do compare their experience of depression and anxiety symptoms to other people's experiences of these

symptoms when judging both how severe their own symptoms are and how worried they are about the occurrence of these symptoms. In all four models reported, how participants' believed their symptom occurrence compared to that of other people in the general population significantly predicted symptom severity and worry judgements. It was the social comparison variable measuring where participants believed their symptom occurrence ranked in comparison to other people in the general population that significantly predicted worry and symptom severity judgements in all four models. This variable accounted for, on average, across both symptoms, 14% of the variance in symptom severity judgements and 10% of the variance in worry judgements. Although the variable measuring how much participants' symptom occurrence differed from what they believed the average symptom occurrence of others in the population to be significantly predicted anxiety symptom severity judgements and worry about depression symptom occurrence, the rank variables were much stronger predictors of these judgements. Comparison across all four models showed that the step 2a models including the rank variables accounted for significantly more variance in symptom severity and worry judgements than the step 2b models. The distance from the average variable accounted for, on average, across both symptoms, less than 1% of the variance in worry and symptom severity judgements. Furthermore, where participants believed their symptom occurrence ranked in comparison to others had a greater bearing on severity and worry judgements than their actual symptom occurrence.

Aim 3

The third aim of the study was to test for direct and indirect associations between beliefs about how symptoms compare to others and help-seeking as outlined in Figure 3.1. Direct associations will be explored first. Tables 3.7 to 3.10 below show the results from regression analyses investigating predictors of:

1. The number of help-seeking behaviours participants engaged in to manage depression (*M* = 1.47, *SD* = 1.63) and anxiety (*M* = 1.20, *SD* = 1.43) symptoms,
2. Whether they had sought help at all for symptoms (depression: 61.7%; anxiety: 57.5%),
3. Whether they had consulted a professional about their symptoms (depression: 26.1%; anxiety: 31.7%), and
4. Whether they had taken medication for symptoms (depression: 22.7%; anxiety: 23.6%).

Total number of help-seeking behaviours: Table 3.7 shows a very similar pattern of results to those reported thus far. The step 1 models each account for 20% of the variance in the total number of help-seeking behaviours participants engaged with in order to manage their symptoms of depression and anxiety with symptom occurrence being the strongest predictor of this. The rank variables account for an additional 5% of explained variance (both significant increases) and when they are added in step 2a the effect of symptom occurrence is attenuated by around 60% in both models making rank the strongest predictor in the models (the rank beta weights are significantly larger than those of other predictors in the models). Although distance from the average is also a significant predictor of the total number of help-seeking behaviours for both depression and anxiety symptoms, compared to symptom occurrence it is a significantly weaker predictor (depression: symptom occurrence β = .33, distance from the average β = .12; anxiety: symptom occurrence β = .32, distance from the average β = .14). Addition of the distance from the average variables resulted in a 1% increase in explained variance (significant for both models), however, the step 2a models including rank accounted for significantly more variance in the number of help-seeking behaviours than the step 2b models including distance from the average, depression: $t(640) = 2.77$, $p = .003$, anxiety: $t(640) = 2.47$, $p = .007$.

Table 3.7.

Multiple Linear Regression Analyses Predicting the Total Number of Help-seeking Behaviours for Each Symptom from Age, Gender, Symptom Occurrence, Direct Rank of Symptom Occurrence, and Direct Distance from the Believed Average Symptom Occurrence of People in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Depression					Anxiety				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.20***				**<.001**	.20***				**<.001**
Constant		1.27 (0.86-1.72)	0.22		.001		0.89 (0.55-1.23)	0.19		.001
Age		**-0.01 (-0.02-0.00)**	**0.00**	**-.07**	**.036**		-0.01 (-0.01-0.00)	0.00	-.05	.133
Gender		**-0.32 (-0.55--0.08)**	**0.12**	**-.10**	**.008**		**-0.26 (-0.48--0.03)**	**0.10**	**-.09**	**.014**
Symptom Occurrence		**0.08 (0.06-0.10)**	**0.01**	**.42**	**.001**		**0.07 (0.05-0.08)**	**0.01**	**.42**	**.001**
Step 2a	.05***				**<.001**	.05***				**<.001**
Constant		1.08 (0.70-1.52)	0.22		.001		0.70 (0.36-1.06)	0.18		.001
Age		-0.01 (-0.02-0.00)	0.00	-.06	.082		0.00 (-0.01-0.00)	0.00	-.03	.366
Gender		**-0.31 (-0.53--0.08)**	**0.11**	**-.09**	**.009**		**-0.28 (-0.50--0.06)**	**0.10**	**-.10**	**.010**
Symptom Occurrence		**0.02 (0.00-0.05)**	**0.01**	**.13**	**.037**		**0.03 (0.00-0.05)**	**0.01**	**.16**	**.012**
Rank		**0.02 (0.01-0.02)**	**0.00**	**.37**	**.001**		**0.02 (0.01-0.02)**	**0.00**	**.34**	**.001**
Step 2b	.01*				**<.001**	.01*				**<.001**
Constant		1.42 (0.96-1.90)	0.23		.001		1.05 (0.66-1.49)	0.21		.001
Age		**-0.01 (-0.02-0.00)**	**0.01**	**-.07**	**.033**		-0.01 (-0.01-0.00)	0.00	-.05	.145
Gender		**-0.33 (-0.56--0.10)**	**0.12**	**-.10**	**.006**		**-0.26 (-0.46--0.08)**	**0.10**	**-.09**	**.015**
Symptom Occurrence		**0.06 (0.04-0.08)**	**0.01**	**.33**	**.001**		**0.05 (0.03-0.07)**	**0.01**	**.32**	**.001**
Distance from the Average		**0.02 (0.00-0.04)**	**0.01**	**.12**	**.037**		**0.02 (0.01-0.04)**	**0.01**	**.14**	**.014**

Note. N = 643, CI = confidence interval. Significant models and predictors are highlighted in bold. *p < .05, ***p < .001.

Table 3.8.

Binary Regression Analyses Predicting Whether Participants Sought Help for Symptoms from Age, Gender, Symptom Occurrence, Direct Rank of Symptom Occurrence, and Direct Distance from the Believed Average Symptom Occurrence of People in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Depression						Anxiety					
Predictors	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.21***				**<.001**		.19***				**<.001**	
Constant			-0.04 (-0.66-0.60)	0.32	.902	0.96			-0.25 (-0.87-0.40)	0.31	.403	0.78
Age		-0.12	-0.01 (-0.03-0.00)	0.01	.071	0.99 (0.97-1.00)		-0.12	-0.01 (-0.03-0.00)	0.01	.120	0.99 (0.97-1.00)
Gender		0.16	0.31 (-0.04-0.64)	0.17	.082	1.36 (0.96-1.92)		0.12	0.23 (-0.14-0.61)	0.18	.206	1.25 (0.89-1.76)
Symptom Occurrence		**0.94**	**0.11 (0.09-0.16)**	**0.02**	**.001**	**1.12 (1.09-1.15)**		**0.90**	**0.10 (0.08-0.13)**	**0.01**	**.001**	**1.10 (1.08-1.13)**
Step 2a	.04***				**<.001**		.05***				**<.001**	
Constant			-0.26 (-0.91-0.34)	0.32	.419	0.77			-0.59 (-1.17-0.01)	0.32	.058	0.56
Age		-0.12	-0.01 (-0.03-0.00)	0.01	.119	0.99 (0.97-1.00)		-0.12	-0.01 (-0.02-0.01)	0.01	.276	0.99 (0.98-1.01)
Gender		0.17	0.33 (-0.01-0.67)	0.18	.059	1.39 (0.97-1.98)		0.15	0.29 (-0.07-0.68)	0.18	.106	1.34 (0.95-1.90)
Symptom Occurrence		**0.43**	**0.05 (0.01-0.09)**	**0.02**	**.030**	**1.05 (1.01-1.09)**		**0.36**	**0.04 (0.01-0.07)**	**0.02**	**.030**	**1.04 (1.00-1.07)**
Rank		**0.63**	**0.02 (0.01-0.04)**	**0.01**	**.001**	**1.02 (1.01-1.03)**		**0.64**	**0.02 (0.01-0.03)**	**0.01**	**.001**	**1.02 (1.02-1.03)**
Step 2b	.00				**<.001**		.01				**<.001**	
Constant			-0.06 (-0.71-0.60)	0.34	.851	0.94			-0.07 (-0.71-0.55)	0.34	.837	0.93
Age		-0.12	-0.01 (-0.03-0.00)	0.01	.076	0.99 (0.97-1.00)		-0.12	-0.01 (-0.03-0.00)	0.01	.123	0.99 (0.97-1.00)
Gender		0.15	0.30 (-0.06-0.65)	0.18	.081	1.35 (0.95-1.92)		0.12	0.23 (-0.13-0.61)	0.18	.192	1.25 (0.89-1.77)
Symptom Occurrence		**1.02**	**0.12 (0.08-0.16)**	**0.02**	**.001**	**1.12 (1.08-1.17)**		**0.72**	**0.08 (0.05-0.12)**	**0.02**	**.001**	**1.09 (1.05-1.12)**
Distance from the Average		0.00	0.00 (-0.04-0.03)	0.02	.829	1.00 (0.97-1.03)		0.19	0.02 (-0.01-0.06)	0.02	.149	1.02 (0.99-1.05)

Note. N = 643, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. ***p < .001.

Table 3.9.

Binary Regression Analyses Predicting Whether Participants Sought Help from a Professional (Doctor, Nurse, Pharmacist or Therapist) for Symptoms from Age, Gender, Symptom Occurrence, Direct Rank of Symptom Occurrence, and Direct Distance from the Believed Average Symptom Occurrence of People in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Depression						Anxiety					
Predictors	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.18***				**<.001**		.10***				**<.001**	
Constant			-2.30 (-2.97--1.65)	0.36	.001	0.10			-1.53 (-2.21--0.93)	0.32	.001	0.22
Age		0.00	0.00 (-0.01-0.02)	0.01	.674	1.00 (0.99-1.02)		0.00	0.00 (-0.02-0.01)	0.01	.895	1.00 (0.98-1.01)
Gender		0.15	0.29 (-0.09-0.68)	0.20	.149	1.33 (0.91-1.95)		0.17	0.33 (-0.06-0.75)	0.19	.073	1.39 (0.98-1.98)
Symptom Occurrence		**0.85**	**0.10 (0.08-0.12)**	**0.01**	**.001**	**1.10 (1.08-1.12)**		**0.54**	**0.06 (0.04-0.08)**	**0.01**	**.001**	**1.06 (1.04-1.08)**
Step 2a	.06***				**<.001**		.04***				**<.001**	
Constant			-2.62 (-3.31--1.99)	0.37	.001	0.07			-1.85 (-2.60--1.22)	0.33	.001	0.16
Age		0.12	0.01 (-0.01-0.02)	0.01	.482	1.01 (0.99-1.02)		0.00	0.00 (-0.01-0.02)	0.01	.719	1.00 (0.99-1.02)
Gender		0.13	0.25 (-0.16-0.68)	0.20	.224	1.28 (0.86-1.90)		0.19	0.37 (-0.02-0.79)	0.19	.052	1.45 (1.01-2.07)
Symptom Occurrence		0.17	0.02 (-0.01-0.06)	0.02	.206	1.02 (0.99-1.06)		0.09	0.01 (-0.02-0.04)	0.02	.672	1.01 (0.98-1.04)
Rank		**0.95**	**0.03 (0.02-0.04)**	**0.01**	**.001**	**1.03 (1.02-1.04)**		**0.64**	**0.02 (0.01-0.03)**	**0.00**	**.001**	**1.02 (1.01-1.03)**
Step 2b	.00				**<.001**		.00				**<.001**	
Constant			-2.20 (-2.95--1.51)	0.36	.001	0.11			-1.55 (-2.24--0.85)	0.34	.001	0.21
Age		0.00	0.00 (-0.01-0.02)	0.01	.679	1.00 (0.99-1.02)		0.00	0.00 (-0.02-0.01)	0.01	.900	1.00 (0.98-1.01)
Gender		0.15	0.29 (-0.08-0.70)	0.20	.135	1.34 (0.91-1.96)		0.17	0.33 (0.02-0.65)	0.18	.070	1.39 (0.98-1.98)
Symptom Occurrence		**0.68**	**0.08 (0.05-0.12)**	**0.02**	**.001**	**1.09 (1.05-1.12)**		**0.54**	**0.06 (0.03-0.10)**	**0.02**	**.001**	**1.06 (1.03-1.10)**
Distance from the Average		0.18	0.02 (-0.02-0.05)	0.02	.341	1.02 (0.98-1.05)		0.00	0.00 (-0.03-0.02)	0.01	.852	1.00 (0.97-1.03)

Note. N = 643, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. ****p* < .001.

Table 3.10.

Binary Regression Analyses Predicting Whether Participants Took Medication (Complementary or Conventional) for Symptoms from Age, Gender, Symptom Occurrence, Direct Rank of Symptom Occurrence, and Direct Distance from the Believed Average Symptom Occurrence of People in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Depression						Anxiety					
Predictors	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.17***				**<.001**		.16***				**<.001**	
Constant			-3.07 (-3.82--2.36)	0.36	.001	0.05			-2.67 (-3.33--2.08)	0.35	.001	0.07
Age		**0.23**	**0.02 (0.00-0.03)**	**0.01**	**.023**	**1.02 (1.00-1.04)**		0.12	0.01 (-0.01-0.03)	0.01	.130	1.01 (1.00-1.03)
Gender		**0.27**	**0.53 (0.13-0.91)**	**0.21**	**.009**	**1.70 (1.13-2.54)**		0.12	0.23 (-0.18-0.65)	0.21	.261	1.26 (0.85-1.86)
Symptom Occurrence		**0.77**	**0.09 (0.07-0.11)**	**0.01**	**.001**	**1.09 (1.07-1.12)**		**0.72**	**0.08 (0.06-0.11)**	**0.01**	**.001**	**1.09 (1.06-1.11)**
Step 2a	.05***				**<.001**		.03***				**<.001**	
Constant			-3.42 (-4.20--2.72)	0.39	.001	0.03			-3.00 (-3.67--2.40)	0.36	.001	0.05
Age		**0.23**	**0.02 (0.00-0.04)**	**0.01**	**.007**	**1.02 (1.00-1.04)**		**0.23**	**0.02 (0.00-0.03)**	**0.01**	**.039**	**1.02 (1.00-1.03)**
Gender		**0.25**	**0.50 (0.09-0.89)**	**0.21**	**.012**	**1.65 (1.09-2.48)**		0.12	0.24 (-0.18-0.68)	0.21	.247	1.28 (0.85-1.91)
Symptom Occurrence		0.09	0.01 (-0.02-0.04)	0.02	.483	1.01 (0.98-1.05)		**0.36**	**0.04 (0.00-0.07)**	**0.02**	**.030**	**1.04 (1.00-1.07)**
Rank		**0.95**	**0.03 (0.02-0.04)**	**0.01**	**.001**	**1.03 (1.02-1.04)**		**0.64**	**0.02 (0.01-0.03)**	**0.01**	**.001**	**1.02 (1.01-1.03)**
Step 2b	.03***				**<.001**		.01				**<.001**	
Constant			-2.68 (-3.53--1.90)	0.38	.001	0.07			-2.49 (-3.32--1.78)	0.39	.001	0.08
Age		**0.23**	**0.02 (0.00-0.04)**	**0.01**	**.031**	**1.02 (1.00-1.04)**		0.12	0.01 (-0.01-0.03)	0.01	.155	1.01 (1.00-1.03)
Gender		**0.28**	**0.55 (0.13-1.01)**	**0.22**	**.011**	**1.74 (1.16-2.62)**		0.11	0.22 (-0.18-0.67)	0.20	.266	1.25 (0.84-1.86)
Symptom Occurrence		0.26	0.03 (-0.01-0.07)	0.02	.077	1.03 (1.00-1.07)		**0.54**	**0.06 (0.03-0.10)**	**0.02**	**.001**	**1.07 (1.03-1.10)**
Distance from the Average		**0.62**	**0.07 (0.03-0.12)**	**0.02**	**.001**	**1.07 (1.03-1.11)**		0.19	0.02 (-0.01-0.06)	0.02	.139	1.02 (0.99-1.06)

Note. N = 643, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. ***p < .001.

Binary logistic regression was used to investigate whether, and (if so) how, people compare to others when making judgements about whether to seek any kind of help for their symptoms, whether to consult a professional (doctor, nurse, pharmacist, or therapist) for their symptoms or whether to take medication (conventional or complementary) for their symptoms. Again, the results were largely consistent with previous findings.

Whether sought help at all: The step 1 models in Table 3.8 show that symptom occurrence was the most important predictor of whether participants sought help for depression and anxiety symptoms - these models account for 19% and 21% of the variance in these behaviours respectively. The addition of the rank variables to the depression and anxiety models resulted in a 4% and 5% increase (both significant) in explained variance respectively and rank becomes the most important predictor of help-seeking. The addition of the distance from the average variables in step 2b had little effect – there was no significant increase in explained variance and these variables were not significant predictors of whether help was sought. Unsurprisingly, model comparison showed that, for both symptoms, there was very strong evidence that the step 2a models fitted the data much better than the step 2b models (differences in -2LL between the step 2a and step 2b models: depression models = -21.26, anxiety models = -25.15).

Whether sought help from a professional: Table 3.9 shows the exact same pattern of results as above apart from that the effect of symptom occurrence is completely removed in the step 2a models. Rank is the most important predictor of whether participants sought help from a professional for depression and anxiety symptoms. Again, there was very strong evidence that the step 2a models fitted the data much better than the step 2b models (differences in -2LL between the step 2a and step 2b models: depression models = -26.71, anxiety models = -21.10). The step 1 models account for 18% and 10% of the variance in professional help-seeking behaviour for depression and anxiety symptoms respectively. Addition of the rank variables resulted in a 6% and 4% increase

(both significant) in explained variance respectively, whereas the addition of the distance from the average variables resulted in less than a 1% increase in both models (both nonsignificant).

Whether took medication: The results presented in Table 3.10 are slightly different. Again, the step 1 models show symptom occurrence to be the most important predictor of whether participants took medication for depression or anxiety symptoms. These models account for 17% and 16% of the variance in these behaviours respectively. When rank is added in step 2a, the effect of symptom occurrence is removed for depression but remains for anxiety, however, rank is the most important predictor in both models. The addition of the rank variables to the depression and anxiety models resulted in a 5% and 3% increase (both significant) in explained variance respectively. The addition of the distance from the average variable in the step 2b anxiety model has no effect. However, when the distance from the average variable is added to the step 1 depression model the effect of symptom occurrence is removed, and the variance accounted for by the model increases significantly by 3%. However, there was very strong evidence that the step 2a depression model fitted the data better than the step 2b depression model (difference in -2LL = -12.60). This was the same for the anxiety models (difference in -2LL between the anxiety step 2a and step 2b model = -13.04).

Summary: Overall, the results provide evidence that people compare their experience of depression and anxiety symptoms to other people's experiences of these symptoms when making decisions about whether to seek help for the symptoms and whether to consult a professional or take medication specifically. In all eight models reported, how participants' believed their symptom occurrence compared to that of other people in the general population significantly predicted these help-seeking outcomes and the number of help-seeking actions that participants took to manage their symptoms. Again, the social comparison variable measuring where participants believed their

symptom occurrence ranked in comparison to other people in the general population significantly predicted all four of the help-seeking outcomes across both symptoms and was the most important predictor of these outcomes. This variable accounted for, on average, across both symptoms, 5% of the variance in the total number of help-seeking behaviours engaged with, decisions about seeking help, and decisions regarding whether to seek help from a professional and 4% of the variance in decisions regarding whether to take medication. In contrast, how participants' symptom occurrence compared to what they believed the average symptom occurrence amongst people in the general population to be was a significant predictor in only three out of the eight models and accounted for, on average, less than 1% of the variance in all help-seeking outcomes. Unsurprisingly, across all help-seeking outcomes, model comparison suggested that there was very strong evidence that the step 2a models including the rank variables fitted the data better than the step 2b models including the distance from the average variables.

Indirect associations: The results so far have shown direct associations between how participants' believe their symptoms of depression and anxiety compare to others and 1) whether or not they think they have depression or anxiety, 2) how worried they are about their experience of these symptoms, 3) how severe they think their symptoms are, and 4) whether they sought help for the symptoms. Previous studies have shown that these aspects of symptom appraisal (beliefs that symptoms represent the presence of a serious underlying disorder, judgements of symptom severity and worry or other emotional responses to symptoms) are also directly linked to seeking help for symptoms (e.g., Elliott et al., 2011; Martin et al., 2003). Therefore, the next section tests whether social comparison also affects help-seeking indirectly through these aspects using mediation analysis. As the strongest direct effects seen were obtained using the variable that measured where participants believed their experience of symptoms ranked in comparison to other people's experiences of these symptoms this variable is used as the measure of

how participants' symptoms compare to others. As in previous analyses age, gender and symptom occurrence are included as covariates in the model. Indirect effects were examined for each aspect separately due to sample size limitations.

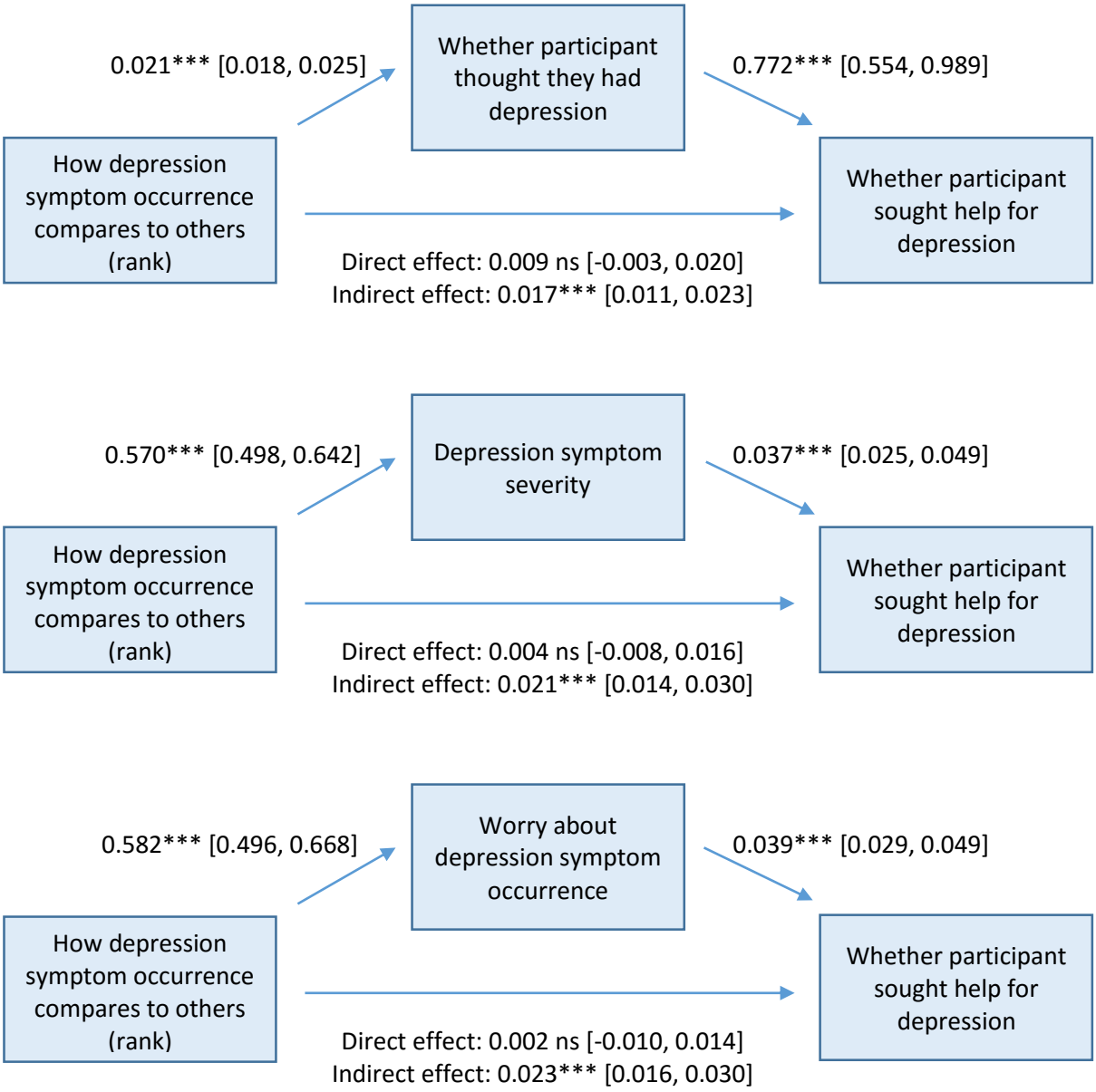

Figure 3.3. Models of rank of depression symptom occurrence as a predictor of help-seeking, mediated by whether participants thought they had depression (top), how severe they thought their symptoms of depression were (middle) and how worried they were about how often they felt depressed (bottom). Figures represent unstandardised regression coefficients with associated 95% confidence intervals (BCa bootstrapped based on 1000 samples) in parentheses. ***$p < .001$, ns = not significant.

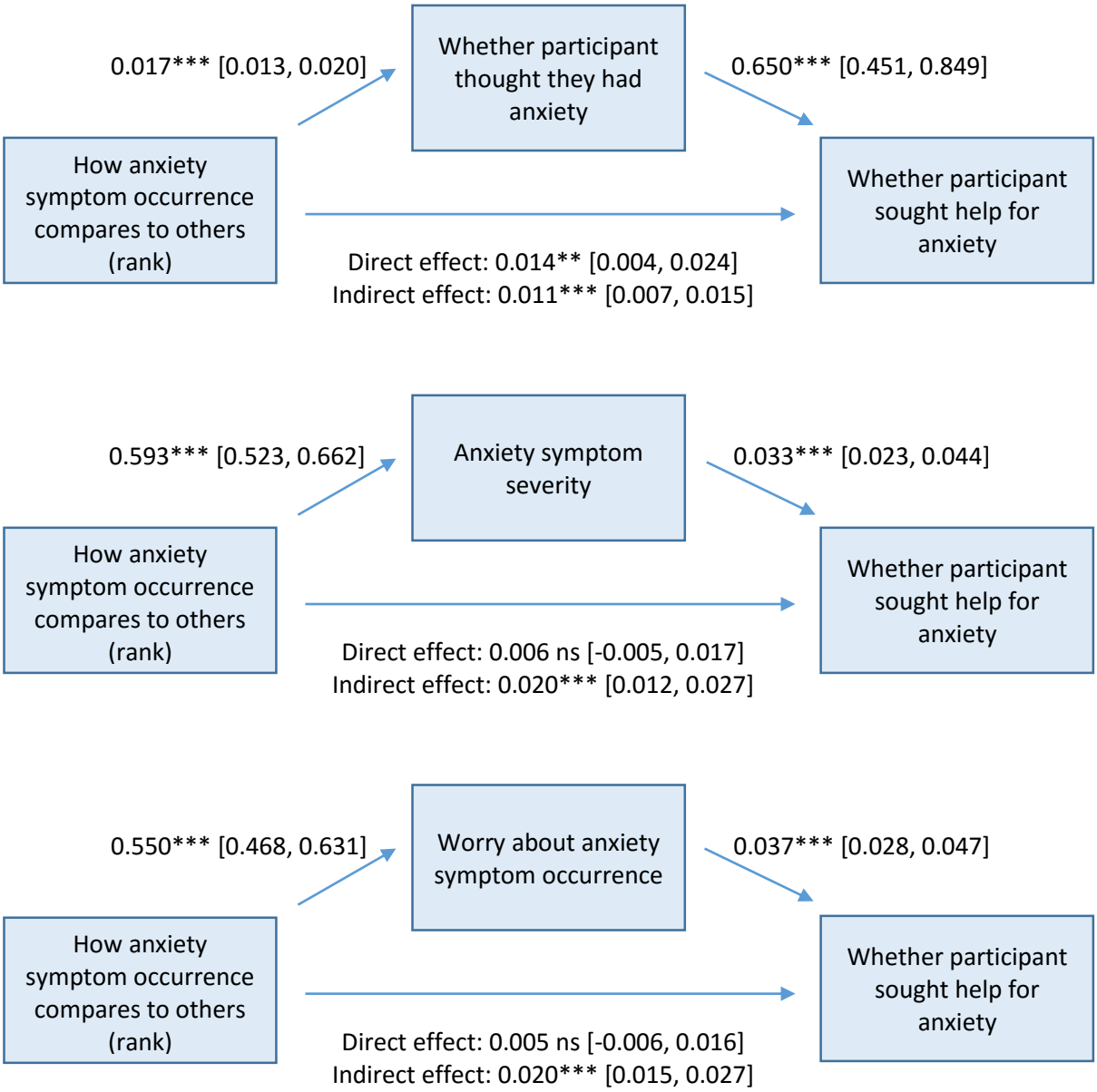

Figure 3.4. Models of rank of anxiety symptom occurrence as a predictor of help-seeking, mediated by whether participants thought they had anxiety (top), how severe they thought their symptoms of anxiety were (middle) and how worried they were about how often they felt anxious (bottom). Figures represent unstandardised regression coefficients with associated 95% confidence intervals (BCa bootstrapped based on 1000 samples) in parentheses. ** $p < .01$, ***$p < .001$, ns = not significant.

Figures 3.3 and 3.4 above show that, for both symptoms, there were significant indirect effects of social comparison on help-seeking through judgements of whether participants thought they had depression/anxiety, symptom severity and worry about symptom occurrence.

Aim 4

The fourth aim of the study was to investigate how people compare to others when making health-related judgements and decisions about help-seeking, namely whether they use rank or average-based comparisons as consistent with DbS and ALT respectively. The results reported above provide strong evidence that when people compare to others to make judgements about whether they have depression or anxiety, how severe their symptoms of depression and anxiety are and how worried they are about these symptoms and when they make decisions about help-seeking, they use their ranked position within the comparison sample (the general population) to do so rather than how much they differ from the sample average.

In the results outlined above, where participants believed their depression and anxiety symptom occurrence ranked in comparison to the frequency with which other people experience these symptoms significantly predicted all seven outcomes across both symptoms. The rank comparison variables were the most important predictors in all of the models where they were included, often attenuating or removing the effect of symptom occurrence. Although the distance from the average variables were significant predictors of the outcomes in half of the models reported, their effect was only greater than that of symptom occurrence in one model.

Comparison of the step 2a (including rank of symptom occurrence) and 2b (including distance from the average symptom occurrence) models across all outcomes and both symptoms consistently showed that the step 2a models accounted for significantly more variance in the outcome than the step 2b models (continuous outcomes) or that there was very strong evidence that the step 2a models fitted the data better than the step 2b models (ordinal and binary outcomes). The rank comparison variables accounted for, on average, across all outcomes and both symptoms, 7% of the variance in the judgements

and decisions measured (range: 3-14%) whereas the distance from the average comparison variable accounted for only 0.5% of the variance, on average (range: 0-3%).

Aim 5

The fifth and final aim of the study was to explore differences in beliefs about others and the potential impact these differences may have on the accuracy of help-seeking decisions.

Beliefs about others: Differences in beliefs about others can be explored using data from the questions asking about the average occurrence of depression and anxiety symptoms in the general population and Figure 3.5 below shows the distribution of answers to these questions. Figure 3.5 shows considerable variation in participants' beliefs about the average occurrence of depression and anxiety in the general population – a comparison group that is the same for all participants (depression: *M* = 8.66, *SD* = 5.84, range = 0-30; anxiety: *M* = 9.72, *SD* = 6.63, range = 0-30). Given such variation, it is likely that many participants will hold incorrect beliefs about the frequency with which other people experience depression and anxiety symptoms. The average number of days a month that participants felt depressed and anxious was 8.92 and 9.69 respectively. If these figures are representative of the general population then this would indicate that the majority of the participants have incorrect beliefs about the occurrence of depression and anxiety symptoms in the U.K.

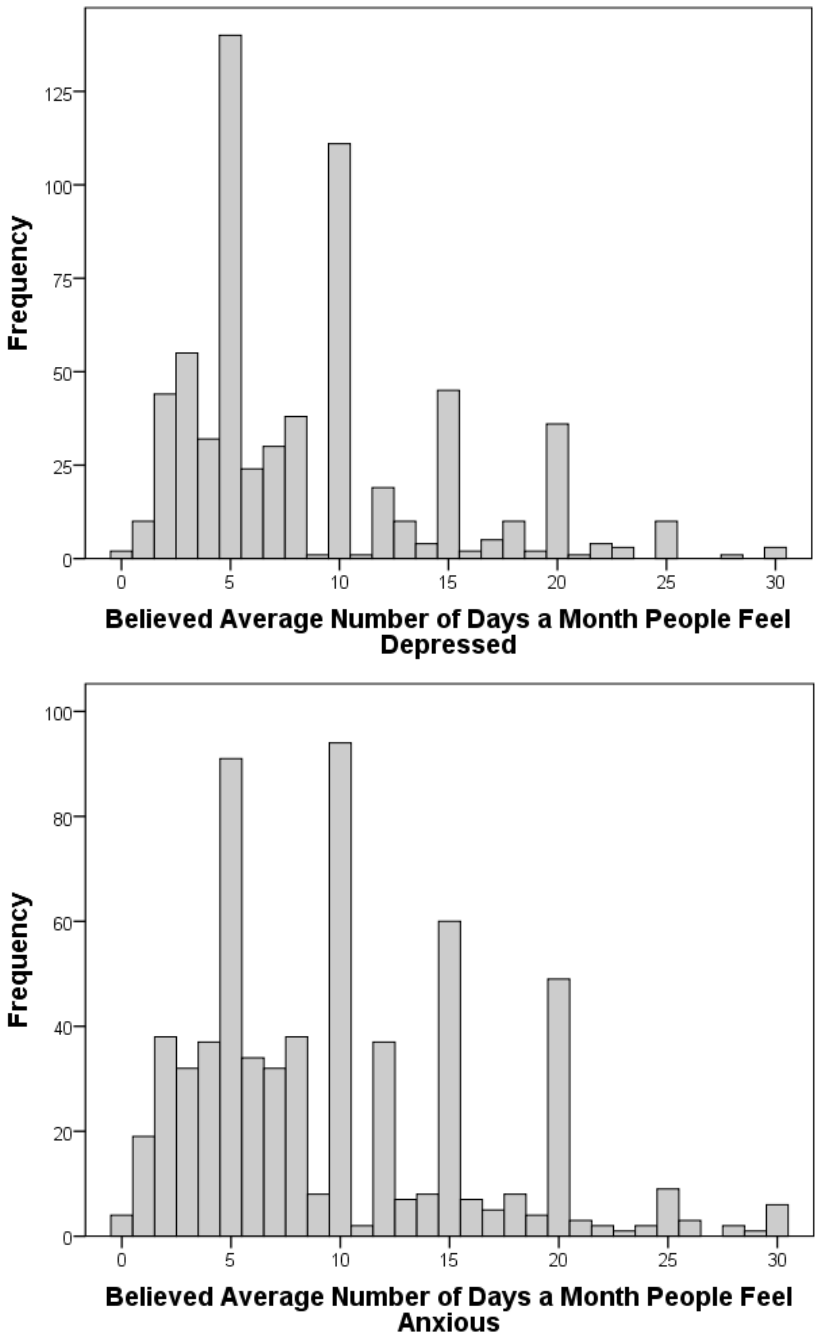

Figure 3.5. Distributions of participants' answers to questions asking them what they thought the average number of days a month people in the general population feel depressed (top) and anxious (bottom).

It is highly likely that a person holding incorrect beliefs about the distribution of symptom occurrence in their comparison sample will make inaccurate judgements about how their symptoms compare to others. Evidence of inaccurate beliefs about where symptoms rank in comparison to others can be seen when looking at the distributions of believed rank amongst participants who experience symptoms for the same number of days a month. Figure 3.6 below gives examples of some of these distributions (which again show variation indicating inaccuracy in some participants) for numbers of days a month symptoms were commonly experienced by participants; 2 (depression *n* = 79, anxiety *n* = 68), 10 (depression *n* = 48, anxiety *n* = 51) and 20 (depression *n* = 48, anxiety *n* = 44) days.

The charts in Figure 3.6 below show that generally these participants ranked themselves appropriately. For the lower and higher days a month that symptoms were experienced the responses are skewed towards the lower and higher ends of the believed rank scale respectively whilst for the middle symptom occurrence of ten days, which was around the average occurrence of all participants, responses are more normally distributed. However, quite a few participants have misjudged their rank. If the number of days a month that the sample experience symptoms of depression and anxiety for is representative of the number of days a month people in the general population experience these symptoms, then 70/74% of people experience symptoms of depression/anxiety respectively on more than 2 days and 11/14% of people experience symptoms of depression/anxiety respectively on more than 20 days. This means that all the participants who experienced depression/anxiety symptoms for two days and thought that they ranked above 30/26 out of 100 were incorrect (depression *n* = 7, anxiety *n* = 7) as were those participants who experienced the symptoms for 20 days and thought they ranked below 89/86 out of 100 (depression *n* = 40, anxiety *n* = 36).

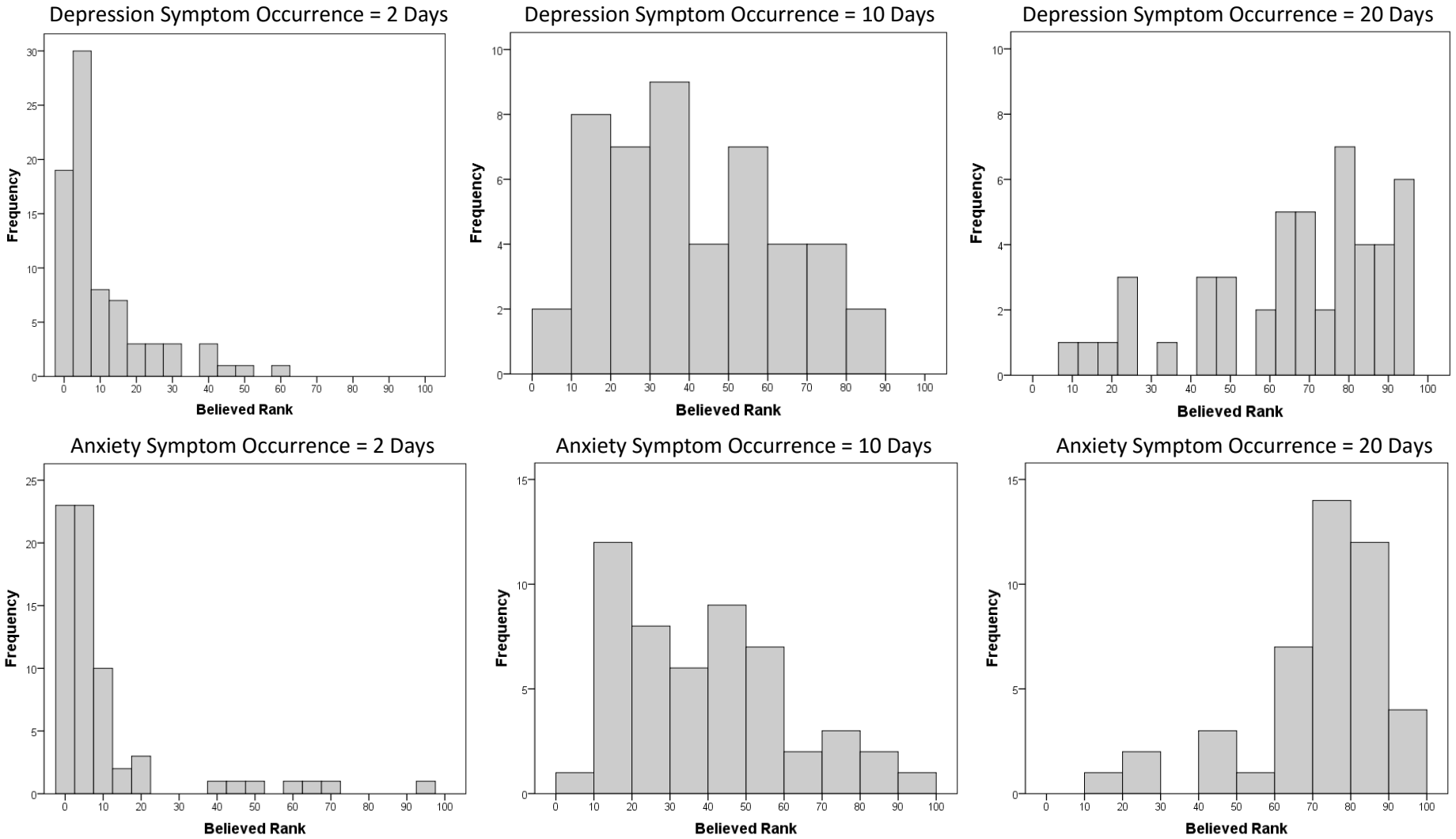

Figure 3.6. Variation in participants' believed rank of their symptom occurrence amongst participants who had the same experience of symptoms.

Help-seeking accuracy: A procedure similar to that used by Mojtabai (2008) was used to investigate help-seeking accuracy. First, participants were split into two groups depending on their rank: those that believed they experienced symptoms of depression/anxiety more frequently than others (i.e., those that gave a rank of 51 or above indicating that they thought 49% or less of the population experienced symptoms on more days than them; depression: n = 188, anxiety: n = 189) and those that believed they experienced symptoms of depression/anxiety less frequently than others (i.e., those that gave a rank of 50 or below; depression: n = 455, anxiety: n = 454). For both depression and anxiety, there was a significant association between beliefs about where participants thought their symptom occurrence ranked in comparison to others and help-seeking (depression: χ^2 (1) = 73.09, p < .001; anxiety: χ^2 (1) = 68.46, p < .001). Participants who thought they experienced symptoms more frequently than others were more likely to seek help than those who thought they experienced symptoms less frequently than others. Of the participants who thought that they experienced symptoms of depression and anxiety more frequently than others, 87.2% (depression) and 82.5% (anxiety) sought help whereas 51.2% (depression) and 47.1% (anxiety) of participants who thought they experienced symptoms less often than others sought help.

In order to investigate help-seeking accuracy, analyses were undertaken again but separately for participants classed as being likely and unlikely to have depression or anxiety based on their PHQ-9 and GAD-7 scores. As a score of 10 or above on both measures indicates the likely presence of clinical levels of depression/anxiety, participants were classed as being likely to have depression or anxiety if they had a score of 10 or above and unlikely to have depression or anxiety if their score was below 10. It should be noted that although both measures have good psychometric properties (as explained earlier), these classifications of “likely” and “unlikely” presence of mental disorder are essentially based on self-report of symptoms and not an actual clinical diagnosis and therefore may not be

completely accurate. As previously mentioned, the PHQ-9 has a sensitivity for major depression of 88% and specificity of 88% and the GAD-7 has a sensitivity for anxiety of 89% and a specificity of 82%. Therefore, the following results should be interpreted as a good estimate of the effect of social comparison on help-seeking accuracy in light of participants' medical information being unavailable/the inability to clinically screen participants.

Again, for both participants who were likely (n = 255) and unlikely (n = 388) to have depression, there was a significant association between beliefs about where participants thought their symptom occurrence ranked in comparison to others and help-seeking (likely to have depression: χ^2 (1) = 10.37, p = .001; unlikely to have depression: χ^2 (1) = 15.45, p < .001). Participants were more likely to seek help when they did not need to if they thought they experienced symptoms more frequently than others. Of the participants who thought that they experienced depression symptoms more frequently than others but were unlikely to have depression, 78.4% sought help whereas only 44.4% of participants who thought they experienced depression symptoms less often than others and who were unlikely to have depression sought help. Similarly, participants likely to have depression were less likely to seek help if they thought they experienced symptoms less frequently than others. Of the participants who likely had depression but thought they experienced depression symptoms less often than others, 26.0% did not seek help whereas only 10.6% of participants who likely had depression and who thought that they experienced depression symptoms more frequently than others did not seek help.

Participants' beliefs about where their symptom occurrence ranked in comparison to others were significantly associated with help-seeking for participants who were unlikely to have anxiety only (likely to have anxiety: χ^2 (1) = 3.29, p = .07; unlikely to have anxiety: χ^2 (1) = 23.95, p < .001). Again, participants were more likely to seek help when they did not need to if they thought they experienced symptoms more frequently than others. Of the participants who thought that they experienced anxiety symptoms more frequently than

others but were unlikely to have anxiety, 73.9% sought help whereas only 42.0% of participants who thought they experienced anxiety symptoms less frequently than others and were unlikely to have anxiety sought help. Although it was nonsignificant, there was a similar trend in help-seeking amongst participants likely to have anxiety as there was for those likely to have depression: 22.7% of participants who thought they experienced anxiety symptoms less frequently than others did not seek help whereas only 12.5% of participants who thought they experienced anxiety symptoms more frequently than others did not seek help.

These findings suggest that accuracy in help-seeking is somewhat linked to the accuracy of participants' beliefs about where their experience of symptoms ranked in comparison to others. Although participants seemingly made inaccurate decisions about help-seeking (i.e., sought help when they were below the screening threshold for depression or anxiety or did not seek help when they were above this threshold) regardless of where they thought their experience of symptoms ranked in comparison to others, they were more likely to make inaccurate help-seeking decisions if they also made inaccurate rank judgements. Participants unlikely to have depression or anxiety were around four times more likely to seek help if they thought they experienced depression/anxiety symptoms more frequently than others (odds ratios: depression = 4.53, anxiety = 3.93). Participants likely to have depression or anxiety were two to three times more likely to not seek help if they thought they experienced symptoms less frequently than others (odds ratios: depression = 2.94, anxiety = 2.10). The implications of these and all of the other findings in this study will be outlined in the general discussion after Study 2.

Study 2: Physical Health

The main aim of Study 2 was to investigate whether the effects seen in Study 1 extend to judgements and decisions made about physical symptoms. The specific aims of the study were as follows:

1. Test for associations between beliefs about how symptoms compare to others and other constructs related to health help-seeking, namely judgements of symptom severity and worry about symptoms,
2. Test for direct and indirect associations between beliefs about how symptoms compare to others and help-seeking as outlined in Figure 3.1,
3. Investigate how people compare to others when making health-related judgements and decisions about help-seeking, namely whether they use rank or average-based comparisons as consistent with DbS and ALT respectively, and
4. Explore differences in beliefs about others.

To this end, a similar procedure to Study 1 was followed in that participants were asked about their experiences of specific symptoms and how they thought their experience of these symptoms compared to others. Four symptoms were chosen as the focus of the study; 1) feeling tired or run down, 2) headaches, 3) muscle, joint or back pain, and 4) cold symptoms. The first three symptoms were chosen as they were identified as being the three most common symptoms in the U.K. in a recent nationwide survey of nearly 2,500 adults (McAteer, Elliott, & Hannaford, 2011) and so it is therefore likely that participants will have experienced them. Participants were asked about their experiences of cold symptoms as this is one symptom where inappropriate help-seeking often occurs, for example, visiting a GP when they cannot treat colds (Braun et al., 2000).

A limitation of Study 1 was that only one aspect of participants' experiences of symptoms was measured – symptom occurrence. In Study 2 symptom duration was also

measured. For two of the symptoms (headaches and colds), participants were asked not just about how frequently they experienced the symptoms and how they thought their symptom occurrence compared to others but also about the duration of their symptoms and how they thought this duration compared to others. As discussed in Chapter 2, it was possible to ask participants about the frequency with which they experienced four symptoms, the duration that they experienced two of these symptoms for and the associated symptom comparison questions because the new, direct comparison measures were used solely in this study - the distribution elicitation questions were not included.

Method

Participants

The 543 participants that completed this study were recruited through CrowdFlower panels and had a mean age of 36.1 years (*SD* = 11.8, range: 18-74 years), were predominantly White (90%, Indian = 2%, Black = 1%, Chinese = 1%, Other = 6%) and 49% were male. Participants were all resident in the U.K. and took part from towns and cities all over the country. Participants were mainly educated up to university (47%) or post-secondary (i.e., education between 16 and 18 years of age such as A-levels: 33%) level (some high school = 4%, finished school at 16 = 16%, rather not say < 1%). Participants were mainly low earners (< £14,999 = 27%, £15,000-£24,999 = 24%, £25,000-£34,999 = 18%, £35,000-£44,999 = 11%, >£45,000 = 10%, rather not say = 10%) and received $0.75 on completion of the study which took 10-15 minutes to complete.

Design and Procedure

The questionnaire that participants completed in this study was designed and hosted using Qualtrics. Participants were asked the following:

Symptom occurrence*:* Occurrence of the four symptoms was obtained through asking: “in the last three months, on how many days have you felt tired or run down?”, “in the last three months, on how many days have you experienced muscle, joint or back

pain?", "in the last three months, how many headaches have you had?" and "in the last three months, how many colds have you had?". A reporting period of three months was chosen to maximise the likelihood of the participant experiencing the symptom and so to reduce the likelihood of floor effects. Participants were asked to consider 3 months to be 90 days. The wording was changed slightly for headaches and colds to ask about how many of each they had experienced in the last 3 months as this seemed more natural than asking about the number of days they had experienced each of these as one headache or cold could last more than one day.

Symptom duration: Typical duration of headaches and colds was measured using the following questions: "typically, how long do your headaches last each time that you have one? (Please answer in number of hours)" and "typically, how long do your colds last each time that you have one? (Please answer in number of days)".

Rank of symptom occurrence and duration within the general population: A similar methodology to Study 1, but without the distribution elicitation questions, was used to elicit participants' beliefs about where the occurrence and duration of their symptoms ranks in comparison to others. This was done using three sets of two questions. First, participants were asked to imagine that all of the adults in the U.K. were lined up in order of the *number of days* in the last 3 months that they have experienced a) feeling tired or run down and then b) muscle, joint or back pain from those that have experienced these symptoms the least number of days in the last 3 months to those that have experienced them the most number of days in the last 3 months. They were then asked where they thought their experience of these symptoms placed them along each line. Secondly, participants did this task again for headaches and cold symptoms but the wording was slightly different in that they were asked to think about the *number of times* rather than the *number of days*. Finally, participants were asked to do exactly the same thing again but thinking about the duration of their symptoms (the amount of time they lasted) rather than

their frequency (how many they had in the last 3 months). They again rated their experience of the symptoms on lines that went from “people who had the shortest headaches/colds in the last 3 months” to “people who had the longest headaches/colds over the last 3 months”. Participants were asked to consider ‘adults’ to be men and women of about their age as McAteer et al (2011) found age-related differences in the prevalence of two of the symptoms. They found that younger participants were more likely to have felt tired or run down over the previous two weeks than older participants who were more likely to have experienced joint pain than younger participants. These effects remained after controlling for sex, marital status, social support, level of education, housing tenure, employment status, household income, ethnicity, smoking status, and the presence of a chronic condition.

Average symptom occurrence and duration in the general population: Participants were asked what they thought the average occurrence of each symptom was: “what do you think the average number of days that people felt tired or run down was over the last three months?”, “what do you think the average number of days that people experienced muscle, joint or back pain was over the last three months?”, “what do you think the average number of headaches that people experienced was over the last three months?” and “what do you think the average number of colds that people experienced was over the last three months?”. They were also asked what they thought the average duration of headaches and colds were: “when people had headaches over the last three months, how long on average do you think they lasted for? (Please answer in number of hours)” and “when people had colds over the last three months, how long on average do you think they lasted for? (Please answer in number of days)”. They were again asked to consider ‘people’ to be men and women of about their age and for three months to be 90 days.

As in Study 1, the following outcomes were measured:

Symptom severity: Participants answered the questions "how severe are your feelings of being tired or run down typically?", "how severe are your muscle, joint or back pains typically?", "how severe are your headaches typically?" and "how severe are your colds typically?" on a slider scale from "not at all severe" to "extremely severe".

Worry: Participants were asked to rate, using a slider scale from "not at all" to "very much" how worried they were about 1) the number of days in the last 3 months they felt tired or run down/ had muscle, joint or back pain, 2) the number of headaches/colds they have had in the last 3 months, and 3) the length of time their headaches/colds typically last.

Help-seeking: For each symptom participants were asked "which of these actions have you taken over the past three months to manage [symptom], tick all that apply" and were provided with the following list (modified slightly from the mental health study to be more relevant to physical symptoms): looked for information, discussed with other people, phoned NHS 24/NHS Direct/NHS 111, consulted a pharmacist, consulted a nurse or doctor, consulted a therapist, took complimentary medication, took over-the-counter medication, took prescribed medication or none of the above.

For all questions where a slider scale was used Qualtrics records where the participant placed the slider as a number between 0 and 100. Questions were allocated into two blocks of independent (symptom occurrence/duration, rank, and average questions) and dependent variables (all outcome measures) and half the participants saw the independent variable block first whilst half saw the dependent variable block first. Within this random presentation order, the presentation order of each of the sets of independent variable questions was randomised. Demographic characteristics (age, gender, level of education, ethnicity, and income bracket) were collected at the end of the survey after all the above questions were presented.

Results

As explained in Chapter 2, the sample size collected was 100 participants short of the minimum 643 participants recommended by the initial power calculation. However, 543 participants is still a large enough sample size to have a 75% chance of detecting a small effect in a regression analysis with 4 predictor variables and, as the analyses below show, this sample size was large enough to detect social comparison effects. During data screening it became apparent that many participants had not read the headache and cold symptom occurrence questions properly. They did not note that they had to answer in terms of the number of *times* in the last 3 months they had experienced these symptoms and not the number of *days* they had experienced these symptoms as they had done for the previous questions asking about the frequency with which they felt tired or run down and experienced muscle joint or back pain. This was evident as some participants were stating that they had experienced upwards of 15 colds in the last 3 months and 60 headaches etc. Unfortunately, there was no way of knowing which participants had read the question and answered correctly and so these data were not used in the analyses reported below. Therefore, the main independent variables included in the analyses are:

Symptom occurrence: This was the number of days in the last 3 months participants felt tired or had no energy (M = 26.06, SD = 27.10) and the number of days in the last month they experienced muscle, joint or back pain (M = 21.60, SD = 29.13).

Symptom duration: This was the number of hours participants typically experienced headaches for (M = 5.79, SD = 10.87) and the number of days they typically experienced colds for (M = 6.03, SD = 6.15).

Rank of symptom occurrence: This rank measure was a direct measure of where the participants thought their experience of muscle, joint or back pain, and feeling tired or run down ranked in comparison to other people's experience of these symptoms. Where participants placed their symptom occurrence on the line that went from "people who

experience the symptom the least" to "people who experience the symptom the most" corresponds to a rank value between 0 and 100 (tired or run down: *M* = 45.09, *SD* = 29.35, muscle, joint or back pain: *M* = 35.05, *SD* = 29.92). Higher values indicate that the participant thought that they experienced the symptom more frequently than the majority of others.

Rank of symptom duration: Again, this was a direct measure of where the participants thought the length of their headaches and colds ranked in comparison to the amount of time other people experience these symptoms for. Where participants placed their symptom duration on the line that went from "people who had the shortest headaches/colds" to "people who had the longest headaches/colds" corresponds to a rank value between 0 and 100 (headaches: *M* = 28.50, *SD* = 25.45, colds: *M* = 31.01, *SD* = 26.65). Higher values indicate that the participant thought that they experienced the symptom for a longer period of time than the majority of others.

Distance from the average symptom occurrence: Participants' answers to the two questions asking what they thought the average occurrence of feeling tired or run down and experiencing muscle, joint or back pain in people from the general population about their age was were deducted from their answers to the relevant symptom occurrence questions. This produced a direct measure of how much their experience of the symptoms differed from what they believed the average experience of others to be. Both distance from the average symptom occurrence variables have possible values ranging from -90 to 90 (tired or run down: *M* = 6.05, *SD* = 25.66, muscle, joint or back pain: *M* = 4.47, *SD* = 28.22). Negative values indicate that the participant experiences the symptom less often than they believe the average person does and positive values indicate that the participant experiences the symptom more often.

Distance from the average symptom duration: Participants' answers to the two questions asking what they though the average length of time people about their age in the

general population experienced headaches and colds for was were deducted from their answers to the relevant symptom duration questions. This produced a direct measure of how much their symptom duration differed from their believed average duration of others (headaches: *M* = 0.18, *SD* = 11.48, colds: *M* = -0.72, *SD* = 8.73). Again, negative values indicate that the participant experiences the symptom for less time than they believe the average person does and positive values indicate that the participant experiences the symptom longer than they believe others do.

As in Study 1, this results section addresses each of the aims outlined in the introduction to Study 2 on page 93 in turn, although the use of rank or average-based comparison strategies (aim 3) is discussed throughout. Correlations conducted prior to the main analyses showed that none of the continuous independent variables were highly correlated (see Table 3.11). Tolerance values obtained through running multiple linear regression analyses confirmed that collinearity is not problematic in any of the models (range of tolerance values for independent variables: tired or run down: .403-.996; muscle, joint or back pain: .334-.968; headaches: .574-.997; colds: .844-.995).

Table 3.11.

Correlations between Continuous Independent Variables.

	Tired or run down				Muscle, joint or back pain			
	Age	Symptom Occurrence	Rank of Symptom Occurrence	Distance from the Average Symptom Occurrence	Age	Symptom Occurrence	Rank of Symptom Occurrence	Distance from the Average Symptom Occurrence
Age	1	-.014	-.063	.009	1	.284***	.146**	.249***
Symptom Occurrence	-.014	1	.655***	.767***	.284***	1	.625***	.807***
Rank of Symptom Occurrence	-.063	.655***	1	.554***	.146**	.625***	1	.527***
Distance from the Average Symptom Occurrence	.009	.767***	.554***	1	.249***	.807***	.527***	1
	Headaches				Colds			
	Age	Symptom Duration	Rank of Symptom Duration	Distance from the Average Symptom Duration	Age	Symptom Duration	Rank of Symptom Duration	Distance from the Average Symptom Duration
Age	1	-.026	-.131**	.004	1	.001	-.171***	.051
Symptom Duration	-.026	1	.414***	.641***	.001	1	.344***	.302***
Rank of Symptom Duration	-.131**	.414***	1	.361***	-.171***	.344***	1	.167***
Distance from the Average Symptom Duration	.004	.641***	.361***	1	.051	.302***	.167***	1

Note. $N = 543$, $^{**}p < .01$, $^{***}p < .001$ (two-tailed).

Aim 1

The first aim of Study 2 was to test for associations between beliefs about how symptoms compare to others and judgements of symptom severity and worry about symptoms. To recap, these judgements were both measured on a 0 to 100 scale with low values indicating mild severity and worry and high values indicating extreme severity and worry. Severity judgements were general, i.e., participants rated how severe their symptoms were overall (tired run down: *M* =39.01, *SD* = 28.26; muscle, joint and back pain: *M* = 33.06, *SD* = 27.83; headaches: *M* = 31.94, *SD* = 25.95; colds: *M* = 31.85, *SD* = 24.45), whereas worry judgements were specific to the aspect of the symptom experience measured, i.e., worry about symptom occurrence and worry about symptom duration (tired run down occurrence: *M* = 33.38, *SD* = 31.24; muscle, joint and back pain occurrence: *M* = 27.35, *SD* =29.52; headaches duration: *M* = 20.71, *SD* = 26.36; colds duration: *M* =20.82, *SD* = 25.46). Multiple linear regression was used to investigate whether the rank and distance from the average variables predicted these judgements and these analyses are reported in Tables 3.12 and 3.13 (symptom severity) and 3.14 and 3.15 (worry) below.

Table 3.12.

Multiple Linear Regression Analyses Predicting Symptom Severity Ratings from Age, Gender, Symptom Occurrence, Rank of Symptom Occurrence, and Distance from the Believed Average Symptom Occurrence of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Tired or Run Down					Muscle, Joint or Back Pain				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.45***				**<.001**	.38***				**<.001**
Constant		29.21 (22.44-36.14)	3.34		.001		30.44 (23.51-37.39)	3.54		.001
Age		**-0.21 (-0.36--0.08)**	**0.08**	**-.09**	**.010**		**-0.20 (-0.37--0.02)**	**0.09**	**-.09**	**.020**
Gender		-1.59 (-5.19-2.06)	1.87	-.03	.400		**-6.74 (-10.56--2.62)**	**1.86**	**-.12**	**.001**
Symptom Occurrence		**0.70 (0.63-0.77)**	**0.04**	**.67**	**.001**		**0.62 (0.55-0.69)**	**0.04**	**.65**	**.001**
Step 2a	.11***				**<.001**	.16***				**<.001**
Constant		15.73 (9.47-21.98)	3.04		.001		17.66 (11.83-23.18)	3.19		.001
Age		**-0.16 (-0.29--0.02)**	**0.07**	**-.07**	**.027**		**-0.16 (-0.32--0.01)**	**0.07**	**-.07**	**.034**
Gender		-0.79 (-3.93-2.29)	1.62	-.01	.658		**-4.15 (-7.85--0.85)**	**1.64**	**-.08**	**.015**
Symptom Occurrence		**0.40 (0.30-0.51)**	**0.05**	**.38**	**.001**		**0.30 (0.20-0.39)**	**0.05**	**.31**	**.001**
Rank		**0.42 (0.33-0.50)**	**0.04**	**.44**	**.001**		**0.48 (0.40-0.58)**	**0.05**	**.52**	**.001**
Step 2b	.02***				**<.001**	.00				**<.001**
Constant		32.23 (25.55-39.04)	3.24		.001		31.61 (24.63-38.20)	3.47		.001
Age		**-0.22 (-0.36--0.09)**	**0.07**	**-.09**	**.006**		**-0.21 (-0.39--0.04)**	**0.08**	**-.09**	**.016**
Gender		-1.64 (-5.54-1.92)	1.89	-.03	.379		**-6.53 (-10.01--2.59)**	**1.93**	**-.12**	**.001**
Symptom Occurrence		**0.55 (0.43-0.66)**	**0.06**	**.53**	**.001**		**0.55 (0.44-0.66)**	**0.06**	**.57**	**.001**
Distance from the Average		**0.21 (0.09-0.33)**	**0.06**	**.19**	**.003**		0.09 (-0.04-0.21)	0.06	.09	.134

Note. N = 543, CI = confidence interval. Significant models and predictors are highlighted in bold. ****p* < .001.

Table 3.13.

Multiple Linear Regression Analyses Predicting Symptom Severity Ratings from Age, Gender, Symptom Duration, Rank of Symptom Duration, and Distance from the Believed Average Symptom Duration of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Headaches					Colds				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.19***				**<.001**	.11***				**<.001**
Constant		37.44 (30.43-44.38)	3.45		.001		35.23 (27.37-41.94)	3.73		.001
Age		**-0.29 (-0.45--0.14)**	**0.08**	**-.13**	**.001**		**-0.27 (-0.43--0.12)**	**0.08**	**-.13**	**.004**
Gender		-1.29 (-5.15-2.69)	2.05	-.03	.533		-1.13 (-5.22-2.40)	2.04	-.02	.605
Symptom Duration		**1.00 (0.82-1.23)**	**0.10**	**.42**	**.001**		**1.18 (0.77-1.89)**	**0.25**	**.30**	**.001**
Step 2a	.27***				**<.001**	.21***				**<.001**
Constant		18.26 (11.80-24.52)	3.19		.001		17.71 (11.89-23.42)	3.16		.001
Age		-0.14 (-0.29-0.00)	0.07	-.07	.057		-0.10 (-0.24-0.03)	0.07	-.05	.155
Gender		-0.56 (-3.76-2.42)	1.62	-.01	.734		1.12 (-2.52-4.59)	1.79	.02	.538
Symptom Duration		**0.43 (0.28-0.60)**	**0.08**	**.18**	**.001**		**0.48 (0.14-1.01)**	**0.20**	**.12**	**.020**
Rank		**0.59 (0.51-0.66)**	**0.04**	**.57**	**.001**		**0.46 (0.38-0.54)**	**0.04**	**.50**	**.001**
Step 2b	.01*				**<.001**	.00				**<.001**
Constant		38.93 (32.80-45.30)	3.41		.001		36.17 (28.49-43.38)	3.86		.001
Age		**-0.30 (-0.46--0.13)**	**0.08**	**-.14**	**.001**		**-0.28 (-0.46--0.12)**	**0.09**	**-.14**	**.002**
Gender		-2.03 (-5.99-2.04)	2.12	-.04	.344		-1.36 (-5.38-2.56)	2.04	-.03	.512
Symptom Duration		**0.82 (0.59-1.05)**	**0.11**	**.34**	**.001**		**1.10 (0.68-1.80)**	**0.26**	**.28**	**.001**
Distance from the Average		**0.27 (0.07-0.52)**	**0.10**	**.12**	**.010**		0.18 (-0.11-0.47)	0.14	.06	.097

Note. N = 543, CI = confidence interval. Significant models and predictors are highlighted in bold. ****p* < .001.

Table 3.14.

Multiple Linear Regression Analyses Predicting Ratings of Worry about the Number of Times Participants Feel Tired or Run Down and Experience Muscle, Joint or Back Pain (Symptom Occurrence), from Age, Gender, Symptom Occurrence, Rank of Symptom Occurrence, and Distance from the Believed Average Symptom Occurrence of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Tired or Run Down					Muscle, Joint or Back Pain				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.34***				**<.001**	.32***				**<.001**
Constant		26.70 (18.36-35.04)	3.86		.001		25.46 (18.79-32.86)	3.78		.001
Age		**-0.24 (-0.41--0.08)**	**0.09**	**-.09**	**.007**		**-0.24 (-0.43--0.06)**	**0.10**	**-.09**	**.019**
Gender		-4.39 (-8.71--0.03)	2.32	-.07	.059		**-5.10 (-8.89--1.01)**	**2.08**	**-.09**	**.012**
Symptom Occurrence		**0.68 (0.59-0.76)**	**0.05**	**.59**	**.001**		**0.60 (0.51-0.70)**	**0.05**	**.59**	**.001**
Step 2a	.10***				**<.001**	.16***				**<.001**
Constant		12.38 (5.31-20.42)	3.65		.002		12.06 (5.61-18.98)	3.39		.001
Age		**-0.18 (-0.35--0.03)**	**0.08**	**-.07**	**.023**		**-0.19 (-0.36--0.01)**	**0.08**	**-.08**	**.020**
Gender		-3.55 (-7.57-0.15)	2.13	-.06	.092		-2.38 (-5.98-1.32)	1.94	-.04	.209
Symptom Occurrence		**0.36 (0.23-0.48)**	**0.06**	**.31**	**.001**		**0.27 (0.15-0.39)**	**0.06**	**.26**	**.001**
Rank		**0.45 (0.35-0.54)**	**0.05**	**.42**	**.001**		**0.51 (0.41-0.60)**	**0.05**	**.51**	**.001**
Step 2b	.01*				**<.001**	.01				**<.001**
Constant		28.87 (21.28-36.61)	3.76		.001		27.04 (20.08-34.38)	3.84		.001
Age		**-0.25 (-0.40--0.09)**	**0.09**	**-.09**	**.006**		**-0.24 (-0.44--0.04)**	**0.10**	**-.10**	**.017**
Gender		**-4.43 (-8.82--0.04)**	**2.29**	**-.07**	**.050**		**-4.82 (-8.59--0.63)**	**2.03**	**-.08**	**.014**
Symptom Occurrence		**0.57 (0.44-0.71)**	**0.07**	**.49**	**.001**		**0.51 (0.37-0.64)**	**0.07**	**.50**	**.001**
Distance from the Average		**0.15 (0.01-0.29)**	**0.07**	**.12**	**.038**		0.12 (-0.02-0.25)	0.07	.12	.074

Note. N = 543, CI = confidence interval. Significant models and predictors are highlighted in bold. *p < .05, ***p < .001.

Table 3.15.

Multiple Linear Regression Analyses Predicting Ratings of Worry about the Duration of Headaches and Colds from Age, Gender, Symptom Duration, Rank of Symptom Duration, and Distance from the Believed Average Symptom Duration of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Headaches					Colds				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.21***				**<.001**	.11***				**<.001**
Constant		30.21 (22.95-37.75)	3.59		.001		31.20 (23.16-39.19)	4.15		.001
Age		**-0.38 (-0.53--0.23)**	**0.08**	**-.17**	**.001**		**-0.41 (-0.57--0.24)**	**0.09**	**-.19**	**.001**
Gender		-3.34 (-7.43-0.60)	2.07	-.06	.110		-3.88 (-8.02-0.06)	2.08	-.08	.069
Symptom Duration		**1.01 (0.71-1.36)**	**0.15**	**.42**	**.001**		**1.06 (0.65-1.67)**	**0.25**	**.26**	**.001**
Step 2a	.28***				**<.001**	.24***				**<.001**
Constant		10.18 (4.20-16.30)	3.28		.002		11.74 (4.61-18.78)	3.52		.001
Age		**-0.22 (-0.36--0.08)**	**0.07**	**-.10**	**.003**		**-0.22 (-0.36--0.08)**	**0.07**	**-.10**	**.003**
Gender		-2.59 (-5.69-0.53)	1.62	-.05	.122		-1.39 (-5.02-2.09)	1.79	-.03	.429
Symptom Duration		**0.42 (0.17-0.74)**	**0.13**	**.17**	**.002**		0.29 (-0.06-0.76)	0.19	.07	.115
Rank		**0.61 (0.52-0.70)**	**0.05**	**.59**	**.001**		**0.51 (0.43-0.59)**	**0.04**	**.54**	**.001**
Step 2b	.00				**<.001**	.00				**<.001**
Constant		30.69 (23.90-37.75)	3.43		.001		31.36 (23.72-37.96)	4.09		.001
Age		**-0.38 (-0.53--0.23)**	**0.08**	**-.17**	**.001**		**-0.41 (-0.59--0.21)**	**0.09**	**-.19**	**.001**
Gender		-3.58 (-7.38-0.58)	2.00	-.07	.080		-3.92 (-8.30-0.43)	2.12	-.08	.062
Symptom Duration		**0.96 (0.73-1.22)**	**0.13**	**.39**	**.001**		**1.05 (0.66-1.69)**	**0.23**	**.25**	**.001**
Distance from the Average		0.08 (-0.21-0.33)	0.12	.04	.464		0.03 (-0.34-0.28)	0.14	.01	.772

Note. N = 543, CI = confidence interval. Significant models and predictors are highlighted in bold. ****p* < .001.

Symptom severity: ***Symptom occurrence:*** Table 3.12 shows that the number of days participants felt tired or run down and experienced muscle, joint or back pain was the strongest predictor of how severe they thought their experience of these symptoms was. These models accounted for 45% and 38% of the variance in these judgements for feeling tired or run down and muscle, joint or back pain respectively. Addition of the rank variables in the step 2a models increased the amount of explained variance significantly, by 11% and 16% respectively. For both symptoms, addition of these variables led to an attenuation of the effect of symptom occurrence by 45-50%. Where the participants believed their symptom occurrence ranked in comparison to others became the strongest predictor of muscle/joint/back pain severity judgements, over and above symptom occurrence (the rank beta weights were significantly greater than the symptom occurrence beta weights). For tired/run down severity judgements, the effect of symptom occurrence and rank was similar (no significant difference in beta weights). For tired/run down, when distance from the average is added to the original step 1 model in step 2b, the amount of variance in symptom severity judgements accounted for by the model increases significantly but only by 2%. In contrast, for muscle/joint/back pain this increase is less than 1% and nonsignificant. How much participants' symptom occurrence differs from what they believe the average symptom occurrence of others to be was a significant predictor of tired/run down severity judgements only. The effect of distance from the average (β = .19) was much weaker than the effect of symptom occurrence (β = .53) in the step 2b tired/run down model. Unsurprisingly, comparison of the predictive ability of the step 2a and 2b models showed that the step 2a models accounted for significantly more variance in symptom severity judgements than the step 2b models, tired/run down: $t(540) = 4.92$, $p < .001$, muscle/joint/back pain: $t(540) = 6.99$, $p < .001$.

Symptom duration*:* Table 3.13 shows exactly the same pattern of results as Table 3.12. In the step 1 models for both headaches and colds, symptom duration is the strongest

predictor of symptom severity judgements. These models accounted for 19% and 11% of the variance in these judgements for headaches and colds respectively. Addition of the rank variables in the step 2a models increased the amount of explained variance significantly, by 27% and 21% respectively. For both symptoms, addition of these variables led to an attenuation of the effect of symptom duration by 50% and where the participants believed their symptom duration ranked in comparison to others became the strongest predictor of symptom severity judgements, again, over and above symptom duration. For headaches, when distance from the average is added in step 2b, the amount of variance in symptom severity judgements accounted for by the model increases significantly but only by 1%, for colds this increase is less than 1% and nonsignificant. How much participants' symptom duration differs from what they believe the average symptom duration of others to be was a significant predictor of headache severity judgements only. The effect of distance from the average (β = .12) was again significantly weaker than the effect of symptom occurrence (β = .34) in the step 2b headache model. Comparison of the predictive ability of the step 2a and 2b models showed that the step 2a models accounted for significantly more variance in symptom severity judgements than the step 2b models, headaches: $t(540) = 8.75$, $p < .001$, colds: $t(540) = 7.12$, $p < .001$.

Worry: *Symptom occurrence:* This pattern of results was much the same for judgements regarding how worried participants were about the number of days they felt tired or run down and experienced muscle, joint or back pain. Again, the step 1 models in Table 3.14 show that symptom occurrence was the strongest predictor of worry ratings (the symptom occurrence beta weights were significantly larger than those of the other predictors in the models) and these models accounted for 34% and 32% of the variance in tired/run down worry ratings and muscle/joint/back pain worry ratings respectively. Although the effect of symptom occurrence is attenuated by around 50% when the rank variables are entered into the models at step 2a, rank becomes the strongest predictor for

muscle/joint/ back pain worry judgements only (the rank beta weight was significantly larger than those of the other predictors). The effect of rank and symptom occurrence is similar in the tired/run down model (no significant difference in beta weights). The amount of explained variance in tired/run down and muscle/joint/ back pain worry ratings increases by 10% and 16% respectively (both significant) when the rank variables are added at step 2a. In contrast, when the distance from the average variables are added to the step 1 model at step 2b, the amount of explained variance increases by less than 1% in both models (significant for the tired/run down model only). Distance from the average is a significant predictor of tired/run down worry ratings only and is a significantly weaker predictor (β = .12) compared to symptom occurrence (β = .49). The step 2a models accounted for significantly more variance in worry ratings than the step 2b models, tired/run down: $t(540) = 4.68$, $p < .001$, muscle/joint/ back pain: $t(540) = 6.40$, $p < .001$.

Symptom duration: The step 1 models in Table 3.15 show that symptom duration was the strongest predictor of how worried participants were about the number of hours they experienced headaches for (the symptom duration beta weight was significantly larger than those of the other predictors) and that symptom duration and age predicted how worried participants were about the number of days they experienced colds for similarly (no significant difference in beta weights). These models accounted for 21% of the variance in worry about headache duration judgements and 11% of the variance in worry about cold duration judgements. When the rank variables are added in step 2a the effect of symptom duration is completely removed from the cold model and is attenuated by around 55% in the headache model with rank becoming the strongest predictor of worry about headache duration (the rank beta weights were significantly larger than those of the other predictors in the models). The amount of explained variance in headache and cold worry ratings increases dramatically by 28% and 24% respectively (both significant). In contrast, when the distance from the average variables are added at step 2b, the amount of explained

variance increases by less than 1% in both models (nonsignificant in both cases) and neither variable is a significant predictor of worry ratings. Unsurprisingly, the step 2a models accounted for significantly more variance in worry ratings than the step 2b models, headaches: $t(540) = 9.49$, $p < .001$, colds: $t(540) = 7.95$, $p < .001$.

Summary*:* Overall, as in Study 1, the results provide evidence that people do compare their experience of symptoms to other people's experiences of these symptoms when judging both how severe their own symptoms are and how worried they are about these symptoms and that they do this for both mental and physical health symptoms. In all of the eight models reported, how participants' symptoms compared to those of other people significantly predicted worry and symptom severity judgements. The findings show that people compare not just their symptom occurrence to that of others but also the duration of their symptoms and suggest that they do so using rank-based strategies. It was the social comparison variable measuring where participants believed their symptom occurrence and duration ranked in comparison to other people around their age in the general population that significantly predicted worry and symptom severity judgements in all eight models. This variable accounted for, on average, 19% of the variance in symptom severity judgements and 20% of the variance in worry judgements. Furthermore, in the majority of cases, where participants believed their symptom occurrence and duration ranked in comparison to others had more bearing on severity and worry judgements than their actual symptom occurrence and duration. In contrast, how participants' symptom occurrence and duration compared to what they believed the average symptom occurrence and duration to be amongst people their age in the general population was a significant predictor in only three out of the eight models and accounted for, on average, less than 1% of the variance in worry and symptom severity judgements. The effect of the distance from the average comparison variable was never as strong as symptom occurrence or duration in the models.

Aim 2

The second aim of the study was to test for direct and indirect associations between beliefs about how symptoms compare to others and help-seeking as outlined in Figure 3.1. Direct associations will be explored first. Tables 3.16 to 3.23 below show the results from regression analyses investigating predictors of:

1. The number of help-seeking behaviours participants engaged in to manage feeling tired/run down (*M* = 1.30, *SD* = 1.48), muscle/joint/back pain (*M* = 1.15, *SD* = 1.31), headaches (*M* = 1.05, *SD* = 0.92) and cold symptoms (*M* = 1.03, *SD* = 1.10),
2. Whether participants sought help at all for symptoms (tired/run down: 58.7%; muscle/joint/back pain: 62.8%; headaches: 77.3%; colds: 67.0%),
3. Whether participants had consulted a professional about their symptoms (tired/run down: 19.5%; muscle/joint/back pain: 17.3%; headaches: 11.0%; colds: 10.1%), and
4. Whether participants had taken medication for symptoms (tired/run down: 34.4%; muscle/joint/back pain: 47.7%; headaches: 68.7%; colds: 57.8%).

Table 3.16.

Multiple Linear Regression Analyses Predicting the Total Number of Help-seeking Behaviours for Each Symptom from Age, Gender, Symptom Occurrence, Rank of Symptom Occurrence, and Distance from the Believed Average Symptom Occurrence of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Tired or Run Down					Muscle, Joint or Back Pain				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.09***				**<.001**	.13***				**<.001**
Constant		1.56 (1.10-2.02)	0.24		.001		1.28 (0.92-1.62)	0.18		.001
Age		**-0.02 (-0.03--0.01)**	**0.01**	**-.12**	**.002**		**-0.01 (-0.02-0.00)**	**0.00**	**-.10**	**.016**
Gender		-0.19 (-0.41-0.02)	0.12	-.06	.112		**-0.21 (-0.40--0.02)**	**0.10**	**-.08**	**.046**
Symptom Occurrence		**0.02 (0.01-0.02)**	**0.00**	**.28**	**.001**		**0.02 (0.01-0.02)**	**0.00**	**.38**	**.001**
Step 2a	.04***				**<.001**	.05***				**<.001**
Constant		1.12 (0.68-1.56)	0.23		.001		0.94 (0.58-1.29)	0.18		.001
Age		**-0.01 (-0.02-0.00)**	**0.01**	**-.11**	**.005**		**-0.01 (-0.02-0.00)**	**0.00**	**-.09**	**.025**
Gender		-0.16 (-0.37-0.04)	0.12	-.05	.171		-0.14 (-0.34-0.05)	0.10	-.05	.169
Symptom Occurrence		0.01 (0.00-0.01)	0.00	.10	.086		**0.01 (0.00-0.01)**	**0.00**	**.19**	**.003**
Rank		**0.01 (0.01-0.02)**	**0.00**	**.27**	**.001**		**0.01 (0.01-0.02)**	**0.00**	**.30**	**.001**
Step 2b	.00				**<.001**	.00				**<.001**
Constant		1.64 (1.19-2.09)	0.24		.001		1.32 (0.94-1.70)	0.19		.001
Age		**-0.02 (-0.03--0.01)**	**0.01**	**-.13**	**.002**		**-0.01 (-0.02-0.00)**	**0.00**	**-.10**	**.015**
Gender		-0.19 (-0.42-0.02)	0.12	-.06	.115		-0.20 (-0.41-0.00)	0.11	-.08	.060
Symptom Occurrence		**0.01 (0.00-0.02)**	**0.00**	**.20**	**.004**		**0.02 (0.01-0.02)**	**0.00**	**.33**	**.001**
Distance from the Average		0.01 (0.00-0.01)	0.00	.10	.132		0.00 (0.00-0.01)	0.00	.06	.350

Note. N = 543, CI = confidence interval. Significant models and predictors are highlighted in bold. ****p* < .001.

Table 3.17.

Multiple Linear Regression Analyses Predicting the Total Number of Help-seeking Behaviours for Each Symptom from Age, Gender, Symptom Duration, Rank of Symptom Duration, and Distance from the Believed Average Symptom Duration of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Headaches					Colds				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.05***				**<.001**	.04***				**<.001**
Constant		1.43 (1.15-1.72)	0.15		.001		1.37 (1.04-1.74)	0.19		.001
Age		**-0.01 (-0.02--0.01)**	**0.00**	**-.14**	**.001**		**-0.01 (-0.02-0.00)**	**0.01**	**-.14**	**.006**
Gender		-0.10 (-0.25-0.05)	0.08	-.05	.237		-0.09 (-0.28-0.08)	0.09	-.04	.318
Symptom Duration		**0.01 (0.01-0.02)**	**0.00**	**.16**	**.003**		**0.03 (0.01-0.04)**	**0.01**	**.14**	**.002**
Step 2a	.05***				**<.001**	.06***				**<.001**
Constant		1.13 (0.82-1.45)	0.17		.001		0.95 (0.59-1.33)	0.19		.001
Age		**-0.01 (-0.02-0.00)**	**0.00**	**-.11**	**.004**		**-0.01 (-0.02-0.00)**	**0.00**	**-.09**	**.041**
Gender		-0.08 (-0.24-0.06)	0.08	-.05	.271		-0.04 (-0.21-0.13)	0.09	-.02	.664
Symptom Duration		0.01 (0.00-0.01)	0.00	.05	.270		0.01 (0.00-0.02)	0.01	.05	.174
Rank		**0.01 (0.01-0.01)**	**0.00**	**.25**	**.001**		**0.01 (0.01-0.01)**	**0.00**	**.27**	**.001**
Step 2b	.00				**<.001**	.01				**<.001**
Constant		1.46 (1.17-1.77)	0.15		.001		1.32 (0.93-1.72)	0.19		.001
Age		**-0.01 (-0.02--0.01)**	**0.00**	**-.14**	**.002**		**-0.01 (-0.02-0.00)**	**0.01**	**-.13**	**.010**
Gender		-0.11 (-0.25-0.03)	0.07	-.06	.140		-0.08 (-0.25-0.09)	0.09	-.04	.390
Symptom Duration		**0.01 (0.00-0.02)**	**0.01**	**.12**	**.038**		**0.03 (0.02-0.05)**	**0.01**	**.17**	**.003**
Distance from the Average		0.01 (0.00-0.01)	0.00	.07	.190		-0.01 (-0.03-0.00)	0.01	-.08	.113

Note. N = 543, CI = confidence interval. Significant models and predictors are highlighted in bold. ****p* < .001.

Table 3.18.

Binary Regression Analyses Predicting Whether Participants Sought Help for Symptoms from Age, Gender, Symptom Occurrence, Rank of Symptom Occurrence, and Distance from the Believed Average Symptom Occurrence of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Tired or Run Down						Muscle, Joint or Back Pain					
Predictors	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.10***				**<.001**		.13***				**<.001**	
Constant			0.47 (-0.18-1.08)	0.33	.150	1.59			0.14 (-0.57-0.80)	0.33	.679	1.14
Age		**-0.24**	**-0.02 (-0.04--0.01)**	**0.01**	**.004**	**0.98 (0.97-0.99)**		0.00	0.00 (-0.02-0.01)	0.01	.701	1.00 (0.98-1.01
Gender		0.14	0.28 (-0.16-0.68)	0.19	.134	1.32 (0.92-1.91)		0.02	0.03 (-0.34-0.44)	0.19	.871	1.03 (0.71-1.49)
Symptom Occurrence		**0.54**	**0.02 (0.01-0.03)**	**0.00**	**.001**	**1.02 (1.01-1.03)**		**0.87**	**0.03 (0.02-0.05)**	**0.01**	**.001**	**1.03 (1.02-1.04)**
Step 2a	.03***				**<.001**		.07***				**<.001**	
Constant			0.04 (-0.65-0.68)	0.34	.885	1.04			-0.36 (-1.09-0.36)	0.35	.278	0.70
Age		**-0.24**	**-0.02 (-0.04-0.00)**	**0.01**	**.011**	**0.98 (0.97-1.00)**		0.00	0.00 (-0.02-0.02)	0.01	.943	1.00 (0.98-1.02)
Gender		0.13	0.25 (-0.20-0.65)	0.19	.182	1.28 (0.89-1.86)		-0.07	-0.13 (-0.49-0.29)	0.20	.500	0.88 (0.60-1.29)
Symptom Occurrence		**0.27**	**0.01 (0.00-0.02)**	**0.01**	**.041**	**1.01 (1.00-1.02)**		**0.29**	**0.01 (0.00-0.03)**	**0.01**	**.049**	**1.01 (1.00-1.02)**
Rank		**0.59**	**0.02 (0.01-0.02)**	**0.00**	**.001**	**1.02 (1.01-1.02)**		**0.90**	**0.03 (0.02-0.04)**	**0.01**	**.001**	**1.03 (1.02-1.03)**
Step 2b	.00				**<.001**		.00				**<.001**	
Constant			0.52 (-0.12-1.14)	0.32	.086	1.69			0.20 (-0.42-0.86)	0.35	.581	1.22
Age		**-0.24**	**-0.02 (-0.04--0.01)**	**0.01**	**.006**	**0.98 (0.97-0.99)**		0.00	0.00 (-0.02-0.02)	0.01	.671	1.00 (0.98-1.01)
Gender		0.14	0.28 (-0.08-0.69)	0.19	.140	1.32 (0.92-1.91)		0.01	0.02 (-0.34-0.40)	0.18	.901	1.02 (0.71-1.48)
Symptom Occurrence		**0.54**	**0.02 (0.01-0.03)**	**0.01**	**.002**	**1.02 (1.01-1.03)**		**0.87**	**0.03 (0.01-0.04)**	**0.01**	**.001**	**1.03 (1.01-1.04)**
Distance from the Average		0.00	0.00 (-0.01-0.02)	0.01	.495	1.00 (0.99-1.02)		0.00	0.00 (-0.01-0.02)	0.01	.469	1.00 (0.99-1.02)

Note. N = 543, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. ****p* < .001.

Table 3.19.

Binary Regression Analyses Predicting Whether Participants Sought Help for Symptoms from Age, Gender, Symptom Duration, Rank of Symptom Duration, and Distance from the Believed Average Symptom Duration of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Headaches						Colds					
Predictors	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.04**				**.003**		.07***				**<.001**	
Constant			2.05 (1.22-2.84)	0.41	.001	7.79			0.66 (-0.03-1.38)	0.35	.057	1.94
Age		**-0.24**	**-0.02 (-0.04-0.00)**	**0.01**	**.011**	**0.98 (0.96-1.00)**		**-0.24**	**-0.02 (-0.03-0.00)**	**0.01**	**.023**	**0.98 (0.97-1.00)**
Gender		-0.15	-0.29 (-0.69-0.12)	0.21	.155	0.75 (0.50-1.13)		0.09	0.18 (-0.15-0.57)	0.19	.328	1.20 (0.83-1.75)
Symptom Duration		0.33	0.03 (-0.01-0.18)	0.04	.206	1.03 (1.00-1.06)		**0.68**	**0.11 (0.06-0.17)**	**0.03**	**.001**	**1.11 (1.06-1.17)**
Step 2a	.06***				**<.001**		.11***				**<.001**	
Constant			1.42 (0.52-2.26)	0.44	.001	4.12			-0.06 (-0.72-0.62)	0.37	.888	0.94
Age		-0.24	-0.02 (-0.04-0.00)	0.01	.068	0.98 (0.97-1.00)		-0.12	-0.01 (-0.02-0.01)	0.01	.322	0.99 (0.98-1.01)
Gender		-0.17	-0.34 (-0.74-0.09)	0.21	.108	0.71 (0.47-1.08)		-0.01	-0.02 (-0.37-0.37)	0.20	.946	0.99 (0.67-1.46)
Symptom Duration		0.00	0.00 (-0.03-0.13)	0.03	.949	1.00 (0.97-1.03)		**0.31**	**0.05 (0.00-0.10)**	**0.02**	**.049**	**1.05 (1.00-1.10)**
Rank		**0.76**	**0.03 (0.01-0.04)**	**0.01**	**.001**	**1.03 (1.01-1.04)**		**0.80**	**0.03 (0.02-0.04)**	**0.01**	**.001**	**1.03 (1.02-1.04)**
Step 2b	.00				**.006**		.00				**<.001**	
Constant			2.07 (1.38-2.81)	0.40	.001	7.93			0.61 (-0.10-1.30)	0.38	.109	1.84
Age		**-0.24**	**-0.02 (-0.04-0.00)**	**0.01**	**.016**	**0.98 (0.96-1.00)**		**-0.24**	**-0.02 (-0.03-0.00)**	**0.01**	**.036**	**0.98 (0.97-1.00)**
Gender		-0.14	-0.27 (-0.64-0.11)	0.21	.189	0.76 (0.51-1.15)		0.09	0.17 (-0.26-0.58)	0.19	.369	1.19 (0.81-1.73)
Symptom Duration		0.22	0.02 (-0.01-0.21)	0.04	.283	1.02 (0.99-1.06)		**0.74**	**0.12 (0.06-0.19)**	**0.03**	**.001**	**1.12 (1.06-1.19)**
Distance from the Average		0.11	0.01 (-0.03-0.06)	0.02	.590	1.01 (0.98-1.03)		-0.17	-0.02 (-0.05-0.01)	0.02	.174	0.98 (0.95-1.02)

Note. N = 543, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. ***p* < .01, ****p* < .001.

Table 3.20.

Binary Regression Analyses Predicting Whether Participants Sought Help from a Professional (Doctor, Nurse, Pharmacist or Therapist) for Symptoms from Age, Gender, Symptom Occurrence, Rank of Symptom Occurrence, and Distance from the Believed Average Symptom Occurrence of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Tired or Run Down						Muscle, Joint or Back Pain					
Predictors	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.09***				**<.001**		.11***				**<.001**	
Constant			-2.41 (-3.29--1.55)	0.43	.001	0.09			-1.62 (-2.28--0.97)	0.37	.001	0.20
Age		0.12	0.01 (-0.02-0.02)	0.01	.623	1.01 (0.99-1.02)		**-0.24**	**-0.02 (-0.05-0.00)**	**0.01**	**.032**	**0.98 (0.96-1.00)**
Gender		0.19	0.37 (-0.09-0.85)	0.23	.099	1.45 (0.92-2.28)		**0.24**	**0.48 (-0.02-0.97)**	**0.25**	**.044**	**1.62 (1.01-2.62)**
Symptom Occurrence		**0.54**	**0.02 (0.01-0.03)**	**0.00**	**.001**	**1.02 (1.01-1.03)**		**0.58**	**0.02 (0.02-0.03)**	**0.00**	**.001**	**1.02 (1.02-1.03)**
Step 2a	.03***				**<.001**		.05***				**<.001**	
Constant			-3.02 (-3.96--2.14)	0.47	.001	0.05			-2.15 (-2.86--1.48)	0.40	.001	0.12
Age		0.12	0.01 (-0.01-0.03)	0.01	.536	1.01 (0.99-1.03)		**-0.24**	**-0.02 (-0.05-0.00)**	**0.01**	**.035**	**0.98 (0.96-1.00)**
Gender		0.18	0.35 (-0.11-0.83)	0.23	.118	1.42 (0.89-2.25)		0.20	0.40 (-0.12-0.88)	0.25	.096	1.50 (0.92-2.44)
Symptom Occurrence		0.27	0.01 (0.00-0.02)	0.01	.052	1.01 (1.00-1.02)		**0.29**	**0.01 (0.00-0.02)**	**0.01**	**.017**	**1.01 (1.00-1.02)**
Rank		**0.59**	**0.02 (0.01-0.03)**	**0.01**	**.001**	**1.02 (1.01-1.03)**		**0.60**	**0.02 (0.01-0.03)**	**0.01**	**.001**	**1.02 (1.01-1.03)**
Step 2b	.00				**<.001**		.00				**<.001**	
Constant			-2.34 (-3.22--1.48)	0.47	.001	0.10			-1.67 (-2.44--0.93)	0.37	.001	0.19
Age		0.12	0.01 (-0.02-0.03)	0.01	.638	1.01 (0.99-1.02)		**-0.24**	**-0.02 (-0.04-0.00)**	**0.01**	**.023**	**0.98 (0.96-1.00)**
Gender		0.19	0.37 (-0.10-0.87)	0.23	.100	1.45 (0.92-2.29)		**0.25**	**0.49 (0.03-0.99)**	**0.25**	**.040**	**1.64 (1.01-2.64)**
Symptom Occurrence		**0.54**	**0.02 (0.00-0.03)**	**0.01**	**.005**	**1.02 (1.01-1.03)**		**0.87**	**0.03 (0.01-0.04)**	**0.01**	**.001**	**1.03 (1.01-1.04)**
Distance from the Average		0.26	0.01 (-0.01-0.02)	0.01	.453	1.01 (0.99-1.02)		0.00	0.00 (-0.02-0.01)	0.01	.676	1.00 (0.98-1.01)

Note. N = 543, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. ****p* < .001.

Table 3.21.

Binary Regression Analyses Predicting Whether Participants Sought Help from a Professional (Doctor, Nurse, Pharmacist or Therapist) for Symptoms from Age, Gender, Symptom Duration, Rank of Symptom Duration, and Distance from the Believed Average Symptom Duration of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Headaches						Colds					
Predictors	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.08***				**<.001**		.05**				**<.001**	
Constant			-1.12 (-2.15--0.09)	0.49	.020	0.33			-1.60 (-2.69--0.50)	0.57	.003	0.20
Age		**-0.47**	**-0.04 (-0.07--0.02)**	**0.01**	**.001**	**0.96 (0.93-0.98)**		**-0.35**	**-0.03 (-0.07--0.01)**	**0.02**	**.041**	**0.97 (0.94-0.99)**
Gender		**0.34**	**0.68 (0.18-1.26)**	**0.30**	**.013**	**1.97 (1.12-3.47)**		**0.33**	**0.66 (0.06-1.28)**	**0.31**	**.028**	**1.93 (1.08-3.45)**
Symptom Duration		0.22	0.02 (-0.01-0.04)	0.01	.065	1.02 (1.00-1.04)		0.18	0.03 (-0.02-0.06)	0.02	.100	1.03 (0.99-1.06)
Step 2a	.02*				**<.001**		.04**				**<.001**	
Constant			-1.62 (-2.74--0.51)	0.54	.001	0.20			-2.35 (-3.57--1.21)	0.61	.001	0.10
Age		**-0.47**	**-0.04 (-0.07--0.02)**	**0.01**	**.001**	**0.96 (0.94-0.99)**		-0.35	-0.03 (-0.06-0.00)	0.02	.097	0.98 (0.95-1.00)
Gender		**0.34**	**0.67 (0.15-1.27)**	**0.30**	**.014**	**1.96 (1.10-3.47)**		**0.30**	**0.60 (0.00-1.19)**	**0.31**	**.047**	**1.82 (1.01-3.28)**
Symptom Duration		0.11	0.01 (-0.02-0.03)	0.01	.500	1.01 (0.99-1.03)		0.00	0.00 (-0.06-0.04)	0.03	.887	1.00 (0.96-1.05)
Rank		**0.25**	**0.01 (0.00-0.03)**	**0.01**	**.009**	**1.01 (1.00-1.03)**		**0.53**	**0.02 (0.01-0.03)**	**0.01**	**.001**	**1.02 (1.01-1.03)**
Step 2b	.00				**<.001**		.02*				**<.001**	
Constant			-1.15 (-2.13-0.01)	0.50	.024	0.32			-1.72 (-2.85--0.57)	0.58	.003	0.18
Age		**-0.47**	**-0.04 (-0.07--0.02)**	**0.02**	**.004**	**0.96 (0.93-0.98)**		-0.35	-0.03 (-0.07-0.00)	0.02	.060	0.97 (0.95-1.00)
Gender		**0.30**	**0.60 (-0.03-1.25)**	**0.30**	**.042**	**1.82 (1.02-3.24)**		**0.31**	**0.61 (0.02-1.21)**	**0.31**	**.039**	**1.84 (1.02-3.32)**
Symptom Duration		**0.33**	**0.03 (0.00-0.06)**	**0.02**	**.028**	**1.03 (1.00-1.06)**		**0.25**	**0.04 (-0.01-0.08)**	**0.02**	**.040**	**1.04 (1.00-1.08)**
Distance from the Average		-0.23	-0.02 (-0.04-0.02)	0.02	.169	0.98 (0.96-1.01)		**-0.17**	**-0.02 (-0.06-0.00)**	**0.02**	**.035**	**0.98 (0.96-1.00)**

Note. N = 543, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. *$p < .05$, **$p < .01$, ***$p < .001$.

Table 3.22.

Binary Regression Analyses Predicting Whether Participants Took Medication (Complementary or Conventional) for Symptoms from Age, Gender, Symptom Occurrence, Rank of Symptom Occurrence, and Distance from the Believed Average Symptom Occurrence of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Tired or Run Down						Muscle, Joint or Back Pain					
Predictors	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.08***				**<.001**		.12***				**<.001**	
Constant			-0.89 (-1.56--0.29)	0.01	.010	0.41			-0.90 (-1.60--0.35)	0.32	.004	0.41
Age		-0.12	-0.01 (-0.02-0.01)	0.19	.473	0.99 (0.98-1.01)		0.24	0.02 (0.00-0.03)	0.01	.061	1.02 (1.00-1.03)
Gender		-0.03	-0.05 (-0.42-0.32)	0.00	.819	0.95 (0.65-1.38)		-0.16	-0.31 (-0.66-0.06)	0.18	.086	0.73 (0.51-1.05)
Symptom Occurrence		**0.54**	**0.02 (0.01-0.03)**	**0.34**	**.001**	**1.02 (1.01-1.03)**		**0.58**	**0.02 (0.01-0.03)**	**0.00**	**.001**	**1.02 (1.01-1.03)**
Step 2a	.03***				**<.001**		03**				**<.001**	
Constant			-1.50 (-2.22--0.85)	0.01	.001	0.22			-1.19 (-1.94--0.57)	0.33	.002	0.31
Age		0.00	0.00(-0.02-0.01)	0.20	.632	1.00 (0.98-1.01)		**0.24**	**0.02 (0.00-0.04)**	**0.01**	**.042**	**1.02 (1.00-1.03)**
Gender		-0.05	-0.09 (-0.48-0.30)	0.01	.669	0.91 (0.62-1.34)		**-0.20**	**-0.39 (-0.74-0.01)**	**0.18**	**.035**	**0.67 (0.47-0.97)**
Symptom Occurrence		0.27	0.01 (0.00-0.01)	0.00	.277	1.01 (1.00-1.01)		**0.29**	**0.01 (0.00-0.02)**	**0.01**	**.019**	**1.01 (1.00-1.02)**
Rank		**0.59**	**0.02 (0.01-0.03)**	**0.37**	**.001**	**1.02 (1.01-1.03)**		**0.30**	**0.01 (0.01-0.02)**	**0.00**	**.001**	**1.01 (1.01-1.02)**
Step 2b	.00				**<.001**		.01				**<.001**	
Constant			-0.76 (-1.49--0.09)	0.01	.030	0.47			-0.78 (-1.44--0.17)	0.33	.014	0.46
Age		-0.12	-0.01 (-0.02-0.01)	0.20	.429	0.99 (0.98-1.01)		0.24	0.02 (0.00-0.03)	0.01	.056	1.02 (1.00-1.03)
Gender		-0.03	-0.05 (-0.44-0.34)	0.01	.796	0.95 (0.65-1.39)		-0.17	-0.33 (-0.71-0.02)	0.18	.066	0.72 (0.50-1.03)
Symptom Occurrence		0.27	0.01 (0.00-0.02)	0.01	.054	1.01 (1.00-1.02)		**0.29**	**0.01 (0.00-0.03)**	**0.01**	**.013**	**1.01 (1.00-1.02)**
Distance from the Average		0.26	0.01 (0.00-0.03)	0.35	.078	1.01 (1.00-1.02)		0.28	0.01 (0.00-0.02)	0.01	.164	1.01 (1.00-1.02)

Note. N = 543, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. ***p* < .01, ****p* < .001.

Table 3.23.

Binary Regression Analyses Predicting Whether Participants Took Medication (Complementary or Conventional) for Symptoms from Age, Gender, Symptom Duration, Rank of Symptom Duration, and Distance from the Believed Average Symptom Duration of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Headaches						Colds					
Predictors	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.04**				**.003**		.02				.071	
Constant			0.81 (0.18-1.49)	0.33	.017	2.24			0.28 (-0.31-0.86)	0.33	.384	1.32
Age		0.12	0.01 (-0.01-0.02)	0.01	.408	1.01 (0.99-1.02)		0.00	0.00 (-0.02-0.01)	0.01	.743	1.00 (0.98-1.010
Gender		**-0.31**	**-0.61 (-0.99--0.24)**	**0.19**	**.002**	**0.54 (0.38-0.79)**		-0.09	-0.18 (-0.56-0.19)	0.18	.340	0.84 (0.59-1.18)
Symptom Duration		0.11	0.01 (-0.01-0.07)	0.02	.379	1.01 (0.99-1.03)		0.25	0.04 (0.00-0.10)	0.02	.062	1.04 (1.00-1.07)
Step 2a	.01**				**<.001**		.08***				**<.001**	
Constant			0.45 (-0.28-1.19)	0.36	.204	1.57			-0.42 (-1.09-0.23)	0.34	.218	0.66
Age		0.12	0.01 (-0.01-0.03)	0.01	.237	1.01 (0.99-1.03)		0.12	0.01 (-0.01-0.02)	0.01	.464	1.01 (0.99-1.02)
Gender		**-0.32**	**-0.63 (-1.03--0.25)**	**0.19**	**.001**	**0.53 (0.37-0.77)**		-0.16	-0.31 (-0.68-0.05)	0.19	.097	0.73 (0.51-1.05)
Symptom Duration		0.00	0.00 (-0.02-0.05)	0.02	.946	1.00 (0.98-1.02)		0.00	0.00 (-0.03-0.05)	0.02	.896	1.00 (0.97-1.04)
Rank		**0.25**	**0.01 (0.00-0.02)**	**0.01**	**.013**	**1.01 (1.00-1.02)**		**0.53**	**0.02 (0.01-0.03)**	**0.00**	**.001**	**1.02 (1.02-1.03)**
Step 2b	.02**				**<.001**		.01				**.034**	
Constant			0.87 (0.23-1.60)	0.34	.009	2.38			0.36 (-0.31-0.99)	0.34	.290	1.43
Age		0.12	0.01 (-0.01-0.02)	0.01	.451	1.01 (0.99-1.02)		0.00	0.00 (-0.02-0.01)	0.01	.694	1.00 (0.98-1.01)
Gender		**-0.27**	**-0.53 (-0.90--0.17)**	**0.20**	**.006**	**0.59 (0.40-0.86)**		-0.08	-0.15 (-0.51-0.19)	0.18	.400	0.86 (0.61-1.22)
Symptom Duration		0.00	0.00 (-0.03-0.05)	0.02	.770	1.00 (0.97-1.02)		0.18	0.03 (-0.01-0.09)	0.02	.189	1.03 (0.99-1.07)
Distance from the Average		**0.34**	**0.03 (0.01-0.08)**	**0.02**	**.018**	**1.03 (1.01-1.06)**		0.17	0.02 (-0.01-0.10)	0.02	.139	1.02 (1.00-1.05)

Note. N = 543, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. ***p* < .01, ****p* < .001.

Total number of help-seeking behaviours: *Symptom occurrence:* Table 3.16 shows a pattern of results very similar to those reported thus far. The step 1 models account for 9% and 13% of the variance in the total number of help-seeking behaviours participants engaged with in order to manage feeling tired/run down and muscle/joint/back pain respectively. Symptom occurrence was the strongest predictor in both models (the symptom occurrence beta weights were significantly larger than those of the other predictors in the models). The rank variables account for an additional 4% (tired/run down model) and 5% (muscle/joint/back pain model) of explained variance (both significant increases). When they are added in step 2a, the effect of symptom occurrence is removed in the tired/run down model and attenuated by around 50% in the muscle/joint/back pain model. Rank was the strongest predictor of the total number of help-seeking behaviours for feeling tired/run down (the rank beta weight was significantly larger than those of the other predictors) but the effect of symptom occurrence (β = .19) and rank (β = .30) on the total number of help-seeking behaviours for muscle/joint/back/pain was not significantly different. When the distance from the average variables are added to the step 1 model in step 2b there is no significant increase in the amount of variance accounted for by either model and these variables are not significant predictors of the total number of help-seeking behaviours. Unsurprisingly, the step 2a models including rank accounted for significantly more variance in the number of help-seeking behaviours than the step 2b models including distance from the average, tired/run down: $t(540) = 2.28$, $p = .012$, muscle/joint/back pain: $t(540) = 2.97$, $p = .002$.

Symptom duration: Table 3.17 shows a similar pattern of results for symptom duration as for symptom occurrence in Table 3.16 apart from the effect of symptom duration, which was completely removed for both models when the rank variables were added in step 2a. The step 1 models account for 5% and 4% of the variance (both significant) in the total number of help-seeking behaviours participants engaged with in

order to manage their headaches and colds respectively, with age and symptom duration being equally strong significant predictors. The rank variables account for an additional 5% (headache model) and 6% (cold model) of explained variance (both significant increases) when they are added in step 2a and become the strongest predictors of the total number of help-seeking behaviours (the rank beta weights were significantly larger than those of the other predictors in the models). Addition of the distance from the average variables in step 2b again results in no significant change in explained variance. Again, the step 2a models including rank accounted for significantly more variance in the number of help-seeking behaviours than the step 2b models which included distance from the average, headaches: $t(540) = 2.88$, $p = .002$, colds: $t(540) = 2.78$, $p = .003$.

Binary logistic regression was used to investigate whether, and (if so) how, people compare to others when making judgements about whether or not to seek any kind of help for their symptoms, whether or not to consult a professional (doctor, nurse, pharmacist or therapist) for their symptoms or whether or not to take medication (over-the-counter, prescription or complementary) for their symptoms. Again, the results were largely consistent with previous findings.

Whether sought help at all: *Symptom occurrence:* The step 1 models in Table 3.18 show that symptom occurrence was the most important predictor of whether participants sought help for feeling tired/run down and muscle/joint/back pain and these models account for 10% and 13% of the variance in these behaviours respectively. The effect of symptom occurrence remains when rank is added in both the step 2a models but rank becomes the most important predictor of help-seeking. The addition of the rank variables to the tired/run down and muscle/joint/back models resulted in a 3% and 7% increase in explained variance respectively (both significant). The step 2b models show that the addition of the distance from the average variables has no effect – these variables are not significant predictors of whether help was sought and add very little additional explained

variance; less than 1% in both models (both nonsignificant increases). Unsurprisingly, model comparison showed that, for both symptoms, there was very strong evidence that the step 2a models fitted the data much better than the step 2b models (differences in -2LL between the step 2a and step 2b models: tired/run down models = -11.86, muscle/joint/back pain models = -32.53).

***Symptom duration**:* Table 3.19 shows that symptom duration was not a significant predictor of whether participants sought help for headaches in any of the models. Age was the sole significant predictor of headache help-seeking in step 1 (this model accounted for 4% of the variance in help-seeking behaviour) but its effect was removed when rank was added to the model in step 2a (resulting in a significant 6% increase in explained variance). Age continues to be the only significant predictor when distance from the average is added in step 2b (resulting in no significant increase in explained variance). Both age and symptom duration significantly predicted whether participants sought help for colds with symptom duration being the most important predictor (this model accounted for 7% of the variance in help-seeking behaviour). When rank is added in step 2a, the effect of age is removed and although the effect of symptom duration remains, rank becomes the most important predictor of whether participants sought help and accounts for additional 11% of variance (significant increase). Again, distance from the average has no effect when it is added in step 2b. Model comparison again showed that, for both symptoms, there was very strong evidence that the step 2a models fitted the data much better than the step 2b models (differences in -2LL between the step 2a and step 2b models: headache models = -22.66, cold models = -47.59).

Whether sought help from a professional: *Symptom occurrence:* Table 3.20 shows symptom occurrence to be the sole significant predictor of whether participants sought help from a professional for feeling tired/run down in step 1. This effect is removed when rank is added to the model in step 2a (rank becomes the only significant predictor in the

model) but symptom occurrence remains the sole significant predictor when distance from the average is added in step 2b. Age, gender and symptom occurrence are all significant predictors of whether participants sought help from a professional for muscle/joint/back pain in the step 1 model. Symptom occurrence is the most important predictor with the effect of age and gender being around the same. When rank is added in step 2a, the effect of gender is removed and rank becomes the most important predictor of whether participants sought help from a professional for muscle/joint/back pain. When distance from the average is added in step 2b, it has no effect and the importance and significance of predictors remains the same as in the step 1 model. The step 1 models account for 9% and 11% of the variance in professional help-seeking behaviour for feeling tired/run down and muscle/joint/back pain respectively. Addition of the rank variables resulted in a 3% and 5% increase in explained variance respectively (both significant) whereas the addition of the distance from the average variables resulted in less than a 1% increase (both nonsignificant). Again, there was very strong evidence that the step 2a models fitted the data much better than the step 2b models (differences in -2LL between the step 2a and step 2b models: tired/run down models = -12.18, anxiety models = -17.64).

Symptom duration: the step 1 headache model in Table 3.21 shows that age and gender were significant predictors of whether participants sought help from a professional for headaches. Age was the most important predictor and the model accounted for 8% of the variance in this behaviour. When rank is added in step 2a, an additional 2% of variance is accounted for by the model (a significant increase). Rank is a significant predictor of seeking help from a professional but age remains the most important predictor. There is no significant increase in explained variance when the distance from the average variable is added in step 2b although the effect of symptom duration becomes significant but again age remains the most important predictor. For colds, age and gender are again significant predictors (with similar effects) of whether participants sought help from a professional for

cold symptoms in step 1. This model accounted for 5% of the variance in this behaviour. The variance accounted for by the model increases by 4% (significant) when rank is added in step 2a, rank becomes a significant predictor and the most important predictor in the model whilst the effect of age is removed. When distance from the average is added in step 2b the variance accounted for by the model increases by 2% (significant) and both distance from the average and symptom duration become significant predictors in the model. The effect of age is removed and gender becomes the most important predictor. This time there was just positive evidence that the step 2a models fitted the data better than the step 2b models (differences in -2LL between the step 2a and step 2b models: headache models = -4.42, cold models = -6.76).

Whether took medication: ***Symptom occurrence:*** The step 1 models presented in Table 3.22 show symptom occurrence to be the sole predictor of whether participants took medication for feeling tired/run down or muscle/joint/back pain. These models account for 8% and 12% of the variance in these behaviours respectively. The addition of the rank variables to the models resulted in a 3% increase (significant) in explained variance in both models. Rank becomes the sole predictor of whether participants took medication for feeling tired/run down, removing the effect of symptom occurrence. In the muscle/joint/back pain step 2a model, all predictors are significant and have similar effects. Addition of the distance from the average variables in the step 2b models does not increase the amount of explained variance significantly. Symptom occurrence remains the sole predictor of whether participants took medication for muscle/joint/back pain but its effect is removed in the tired/run down model. There was very strong evidence that the step 2a tired/run down model fitted the data better than the step 2b model (difference in -2LL = -16.95). There was strong evidence that the step 2a muscle/joint/back pain model fitted the data better than the step 2b model (difference in -2LL = -9.53).

Symptom duration: Table 3.23 shows that symptom duration was not a significant predictor of whether participants took medication for headaches or colds in any of the models. Gender was the sole significant predictor of whether participants took medication for headaches in the step 1 model, which accounted for 4% of the variance in this behaviour. When rank is added in step 2a an additional 1% of variance is accounted for (significant) and rank is a significant predictor of whether participants took medication for headaches but gender remains the most important predictor in the model. When distance from the average is added in step 2b an additional 2% of variance is accounted for (significant) and distance from the average becomes the most important predictor in the model although the effect of gender remains. For colds, the step 1 model does not predict whether participants took medication any better than the baseline model (i.e., the model is not significant). However, when rank is added in step 2a the model becomes significant and rank becomes the sole predictor of whether participants took medication for colds, accounting for 8% of the variance (significant) in this behaviour. The addition of distance from the average to the model has no effect. As these results suggest, this time there is no difference in the fit of the step 2a and 2b headache models (difference in -2LL = 0.98), although there is very strong evidence that the step 2a cold model fitted the data better than the step 2b model (difference in -2LL = -31.89).

Summary: Overall, as in Study 1, the results provide strong evidence that people compare their experience of physical symptoms to other people's experiences of these symptoms when making decisions about whether to seek help for the symptoms and whether to consult a professional or take medication specifically. In all of the 16 models reported above, participants' beliefs about how their symptoms compared to those of other people significantly predicted these help-seeking outcomes and the number of help-seeking actions that participants took to manage their symptoms. Again, the findings show that, when making help-seeking decisions, people compare not just their symptom

occurrence to that of others but also the duration of their symptoms and suggest that they do so using rank-based strategies. Again, the social comparison variable measuring where participants believed their symptom occurrence and duration ranked in comparison to other people around their age in the general population significantly predicted all four of the help-seeking outcomes across all symptoms, and, in the majority of cases, was the most important predictor. This variable accounted for, on average, 5% of the variance in the total number of help-seeking behaviours engaged with, 7% of the variance in decisions about seeking help and 4% of the variance in decisions regarding whether to seek help from a professional and whether to take medication. In contrast, how participants' symptom occurrence and duration compared to what they believed the average symptom occurrence and duration to be amongst people their age in the general population was a significant predictor in only 2 out of the 16 models and accounted for, on average, less than 1% of the variance in all help-seeking outcomes.

Indirect associations: As in Study 1, the results so far have shown direct associations between how participants' believe the frequency with which they experience feeling tired or run down and muscle, joint or back pain and the length of time they experience headaches and colds for compares to other people's experiences of these symptoms and how worried they are about their experience of these symptoms, how severe they think their symptoms are and whether they sought help for the symptoms. As in Study 1, the next section tests whether social comparison also affects help-seeking indirectly through worry about symptoms and judgements of symptom severity using mediation analysis. Again, the rank comparison variable is used in this analysis as the strongest direct effects seen were obtained using this variable. Age, gender and symptom occurrence are included as covariates in the model. Indirect effects were examined for each aspect separately due to sample size limitations. Figures 3.7 and 3.8 below show that, for all symptoms, there

were significant indirect effects of social comparison on help-seeking through judgements of symptom severity and worry about symptom occurrence/duration.

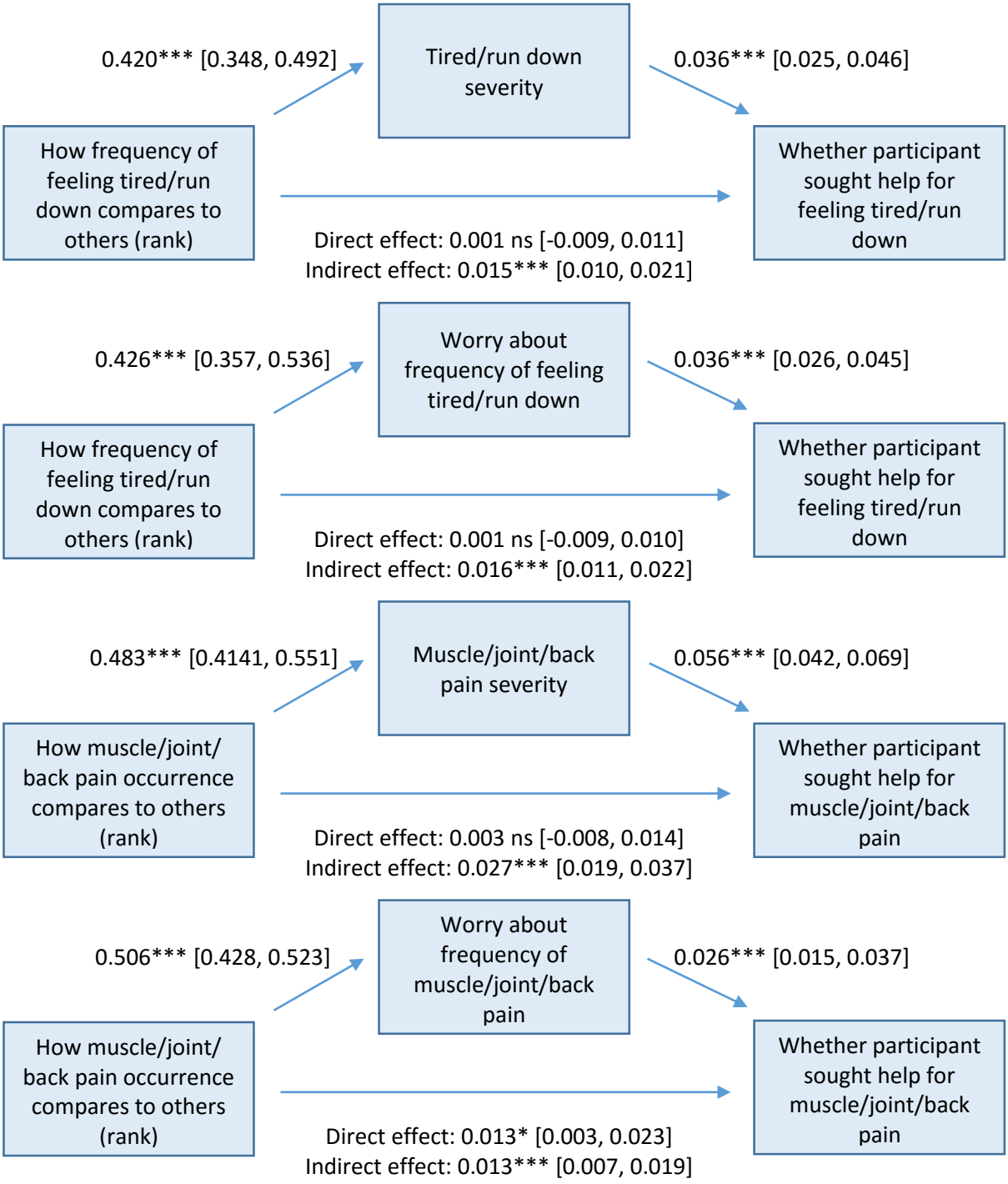

Figure 3.7. Models of rank of symptom occurrence as a predictor of help-seeking, mediated by how severe participants thought their symptoms were and how worried they were about how often they experienced symptoms for feeling tired or run down (top two) and muscle, joint or back pain (bottom two). Figures represent unstandardised regression coefficients with associated 95% confidence intervals (BCa bootstrapped based on 1000 samples) in parentheses. *p < .05, ***p < .001, ns = not significant.

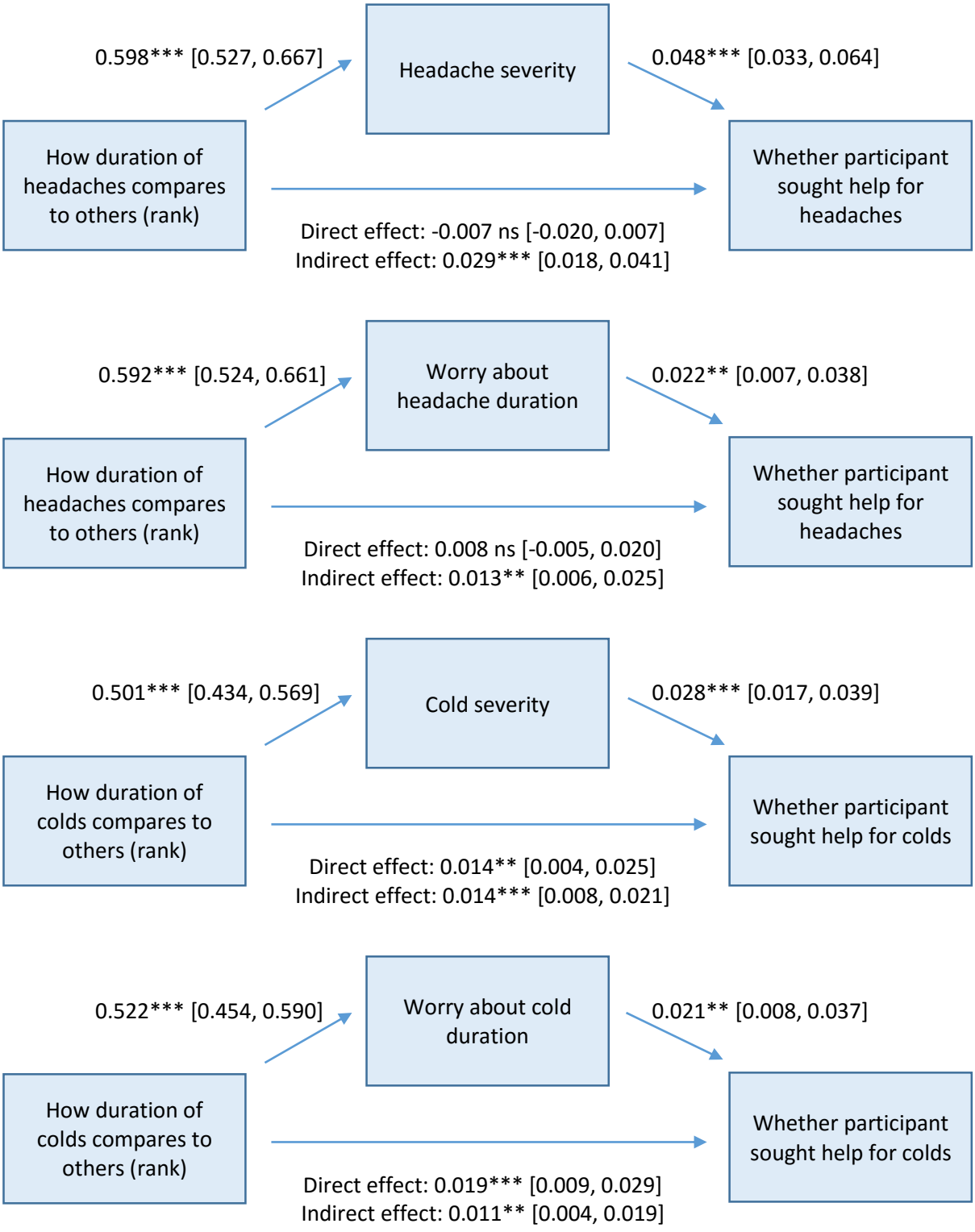

Figure 3.8. Models of rank of symptom duration as a predictor of help-seeking, mediated by how severe participants thought their symptoms were and how worried they were about the length of time that they experienced symptoms for, for headaches (top two) and colds (bottom two). Figures represent unstandardised regression coefficients with associated 95% confidence intervals (BCa bootstrapped based on 1000 samples) in parentheses. ** $p <$ .01, ***$p <$ .001, ns = not significant.

Aim 3

The third aim of the study was to investigate how people compare to others when making health-related judgements and decisions about help-seeking, namely whether they use rank or average-based comparisons as consistent with DbS and ALT respectively. As in Study 1, the results reported above provide strong evidence that when people compare to others to make such judgements and decisions they use their ranked position within the comparison sample (people in the general population about their age) to do so rather than how much they differ from the average of the sample. Across all four symptoms, participants' beliefs about where their experience of symptoms (i.e., their occurrence and duration) ranked in comparison to other people's experience of the same symptoms predicted all six outcome measures (judgements of symptom severity, worry about symptoms, and decisions regarding whether to seek help, how much help to seek, and whether to consult a professional or take medication). In the majority of cases, the rank comparison variable was the most important predictor in the model, often attenuating or removing the effect of participants' absolute experience of the symptom (i.e., their actual symptom occurrence or duration that they experienced the symptom). In contrast, how participants' experiences of the symptoms compared to what they believed the average experience of others to be was a significant predictor in only 5 out of the 24 models and the effect of the distance from the average variable was only greater than the effect of the absolute symptom experience in one of those models.

Comparison of the step 2a (including the rank variables) and 2b (including the distance from the average variables) models across all outcomes and symptoms showed that the step 2a models accounted for significantly more variance in the outcome than the step 2b models (continuous outcomes) or that there was good to very strong evidence that the step 2a models fitted the data better than the step 2b models (ordinal and binary outcomes) for all but one of the 24 models where there was no difference in the fit of the

step 2a compared to the step 2b model. The rank comparison variable accounted for, on average, across all outcomes and symptoms, 10% of the variance in the judgements and decisions measured (range: 1-28%) whereas the distance from the average comparison variable accounted for only 0.5% of the variance, on average (range: 0-2%). Table 3.24 below shows a detailed breakdown of the effects of the two comparison variables across all symptoms and outcomes and both studies.

Table 3.24.
The Amount of Variance (percentage) accounted for by the Rank and Distance from the Average Variables Across all Analyses and the Associated Size of These Effects.

		Symptom					
Outcome	Variable	Depression	Anxiety	Tired or Run Down	Muscle, Joint or Back Pain	Headache	Cold
Thinks Has Disorder	Rank	9[a]	6[a]	-	-	-	-
	DftA	0[a]	1[a]	-	-	-	-
Symptom Severity	Rank	13[a]	14[b]	11[a]	16[b]	27[c]	21[b]
	DftA	0	0[a]	2[a]	0	1[a]	0
Worry	Rank	11[a]	9[a]	10[a]	16[b]	28[c]	24[b]
	DftA	0[a]	0	1[a]	1	0	0
No. of Help-seeking Behaviours	Rank	5[a]	5[a]	4[a]	5[a]	5[a]	6[a]
	DftA	1[a]	1[a]	0	0	0	1
Whether Sought Help	Rank	4[a]	5[a]	3[a]	7[a]	6[a]	11[a]
	DftA	0	1	0	0	0	0
Professional Consulted	Rank	6[a]	4[a]	3[a]	5[a]	2[a]	4[a]
	DftA	0	0	0	0	0	2[a]
Medication Taken	Rank	5[a]	3[a]	3[a]	3[a]	1[a]	8[a]
	DftA	3[a]	1	0	1	2[a]	1

Note. N = 643 for depression and anxiety, *N* = 543 for all other symptoms. DftA = distance from the average. Letters represent effect sizes calculated by transforming the R^2 values into f^2 values; a = small effect, b = medium effect, c = large effect. Where no letter is denoted, this indicates a nonsignficant effect.

Aim 4

The fourth and final aim of the study was to explore differences that participants may have in their beliefs about others. The results so far have shown that people compare to others when making judgements about their health and decisions about health help-seeking. It is hypothesised that people may make inaccurate judgements and decisions when using social comparison to do so if their beliefs about others are inaccurate.

Unfortunately, it was not possible to determine whether participants made accurate help-seeking decisions for the symptoms that were measured in the same way that this was investigated in Study 1. In Study 1 the symptoms measured (feeling depressed and anxious) were core symptoms of mental health disorders (depression and anxiety) and it was possible, using standardised measures, to ascertain whether participants met the clinical threshold on the measures for these disorders and therefore whether their help-seeking was appropriate or not. In Study 2 this was not the case; the symptoms chosen were simply commonly experienced symptoms which may have been experienced in isolation or with other symptoms and so it is difficult to ascertain whether help-seeking was appropriate or not. Therefore, this section will investigate differences in beliefs about others that participants may have which could potentially affect the accuracy of their help-seeking. As in Study 1, these differences can be explored using data from the questions asking about the average occurrence/duration of the symptoms in the general population and where participants believe their experience of the symptoms ranks in comparison to others. As participants were asked to think about people about their age when answering this question (because previous research has shown age-related differences in the frequency with which people experience the symptoms, e.g., older people are more likely to experience joint pain), differences are explored in participants aged 30-40 as this age bracket had the highest number of participants (*n* = 185). Figure 3.9 below shows the distribution of answers to the questions asking what participants thought the average occurrence of feeling tired or run down and muscle, joint or back pain (days in the last 3 months) was in people about their age and what they thought the average duration of headaches (in hours) and colds (in days) was.

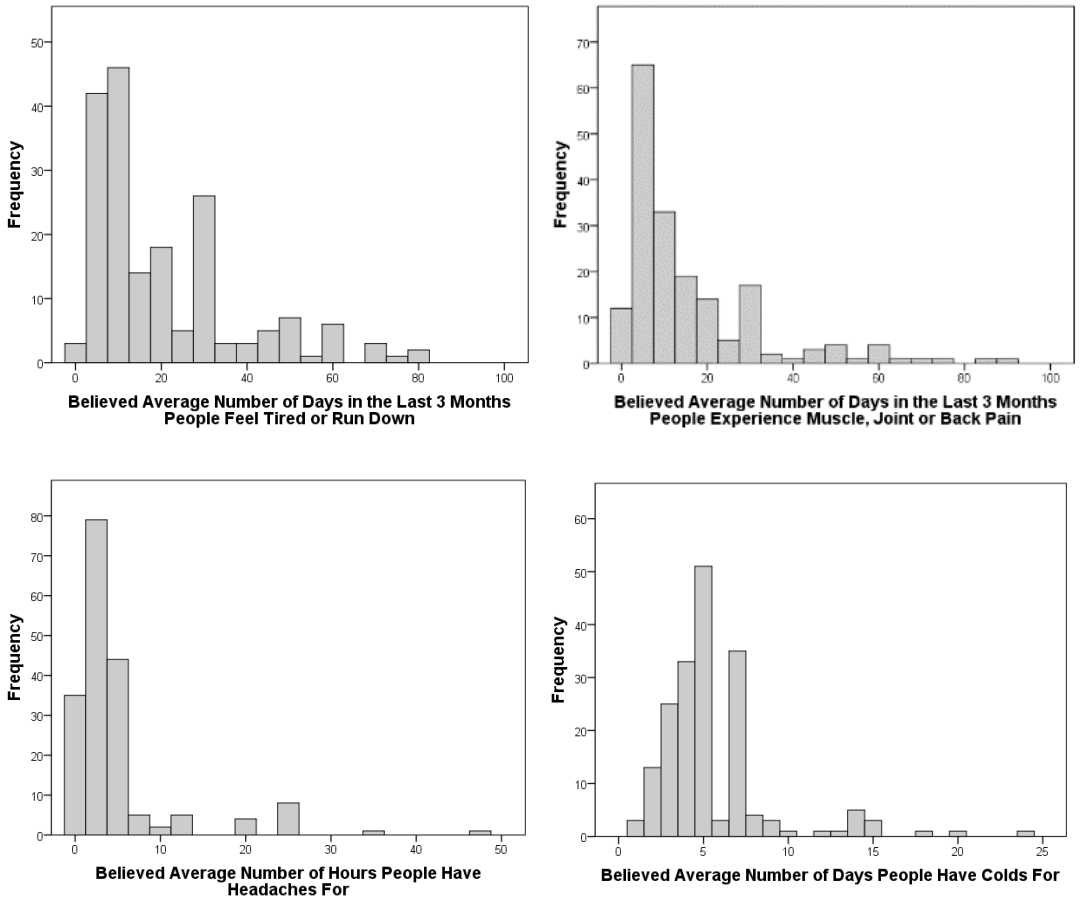

Figure 3.9. Distributions of participants' (aged 30-40) answers to questions asking them what they thought the average number of days in the last three months people about their age in the general population felt tired or run down (top left) and experienced muscle, joint or back pain (top right) and the what they thought the average duration of headaches (bottom left) and colds (bottom right) people about their age in the general population was.

As in Study 1, considerable variation in participants' beliefs about the average occurrence and duration of symptoms existed; tired/run down occurrence: *M* = 20.50, *SD* = 17.83, range = 0-80 days; muscle/joint/back pain occurrence: *M* = 16.14, *SD* = 16.87, range = 0-90 days; headache duration: *M* = 4.92, *SD* = 6.66, range = 0.5-48 hours; cold duration: *M* = 5.57, *SD* = 4.55, range = 1-35 days. Given such variation, it is likely that many participants will hold incorrect beliefs about the frequency and duration with which other people experience symptoms. The average number of days in the last 3 months that participants aged 30-40 felt tired or run down and experienced muscle, joint or back pain was 25.25 and 17.17 respectively. The average amount of time that participants aged 30-40

experienced a headache for was 7.92 hours and a cold for was 1.52 days. If these figures are representative of the general population it would indicate that many of the participants have incorrect beliefs about the occurrence and duration of the symptoms measured in people about their age in the U.K.

It is likely that a person holding incorrect beliefs about the distribution of aspects of symptoms such as their occurrence and duration in their comparison sample will make inaccurate judgements about how their symptoms compare to others. Evidence of inaccurate beliefs about where symptoms rank in comparison to others can be seen when looking at the distributions of believed rank amongst participants who experience symptoms for the same frequency/duration. Figure 3.10 below gives examples of some of these distributions (which again show variation indicating inaccuracy in some participants) for the frequencies/durations with which the symptoms were most commonly experienced by participants aged 30-40; tired/run down: 20 days (*n* = 22); muscle/joint/back pain: 5 days (*n* = 24); headaches: 1 hour (*n* = 45) and colds: 7 days (*n* = 32).

As in Study 1, an idea of the number of participants misjudging their rank can be obtained through estimating what the actual rank of participants' symptom occurrence/duration is and seeing how close this is to the participants' estimates of their rank. This is again done using the distribution of symptom occurrence and duration in participants aged 30-40 and assuming that this is representative of all 30-40 year olds in the population. Thirty eight percent of participants felt tired or run down on *more* than 20 days and fifty percent felt tired or run down on *less* than 20 days. Therefore, participants estimating their rank to be between 50 and 62 (100 minus 38) would be correct. Only 4 out of 22 participants (18%) estimated their rank to be within this bracket. The percentage of participants correctly estimating their rank of muscle/joint/back pain symptom occurrence was 8% (2 out of 24), headache duration was 40% (18 out of 45) and cold duration was 6% (2 out of 32). Although this is quite a crude way of investigating the accuracy of

participants' beliefs about how they rank in comparison to others it does highlight the fact that inaccuracy is common and could affect the accuracy of judgements about symptoms and decisions regarding help-seeking when social comparison is being used to make these judgements and decisions.

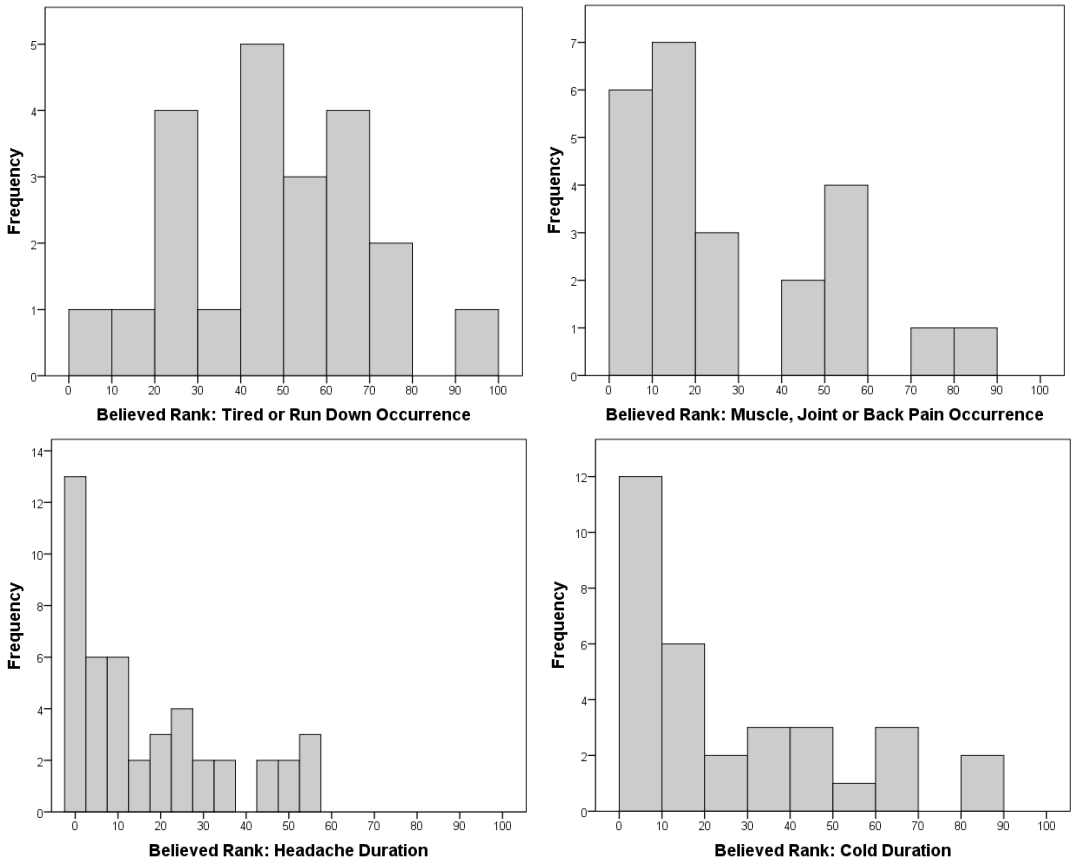

Figure 3.10. Variation in participants' (aged 30-40) believed rank of symptom occurrence/duration amongst participants who had the same experience of symptoms.

Discussion

The two studies reported above aimed to replicate the findings from Melrose et al. (2013) whereby participants' judgements of whether they thought they had depression or anxiety were predicted by their beliefs about where their depression/anxiety symptom occurrence ranked in comparison to others, over and above their actual symptom occurrence. The current studies also aimed to extend these findings in the following ways:

1. By investigating whether participant's beliefs about how their symptoms compare to others are also associated with decisions regarding help-seeking for symptoms both

directly and indirectly through other aspects of symptom appraisal known to affect help-seeking decisions (namely judgements of the presence of an underlying disorder, symptom severity and worry about symptoms),

2. By exploring the above in somatic as well as psychological symptoms and using more than one aspect of symptom experience; occurrence and duration,
3. By further investigating how people compare to others, i.e., whether they use rank or average-based strategies, and
4. By investigating whether comparison of symptoms can lead to inaccurate judgements and decisions.

An overview of the findings relating to these aims will be given first followed by explanations of the implications of the results, suggestions for future research and limitations of the studies.

Findings

Replication*:* The finding of Melrose et al. (2013) that participant's beliefs about how their experience of depression and anxiety symptoms compared to that of others (specifically, where they believed their symptom occurrence ranked in comparison to others) predicted whether they thought they had depression or anxiety was replicated in Study 1 using a much larger sample (*N* = 643 compared to *N* = 133 and 135) and two different methods of eliciting how participants thought their experience of symptoms compared to that of others. The rank comparison variables accounted for 9% and 6% of the variance in judgements of whether participants thought they had depression and anxiety respectively. For depression, the rank variable was the strongest predictor of this judgement, over and above symptom occurrence, as in the original study. For anxiety, the effect of the rank and symptom occurrence variables was comparable.

Help-seeking*:* Help-seeking decisions such as whether participants sought any kind of help at all, the number of help-seeking behaviours engaged with, and whether

professionals were consulted or medication taken for symptoms, were all significantly predicted by where participants believed their experience of symptoms ranked in comparison to other people's experiences of the same symptoms. This was the case across both studies and all six symptoms measured: feeling down, depressed or hopeless; feeling nervous, anxious or on edge; feeling tired or run down; muscle, joint or back pain; headaches and colds. The average amount of variance in these help-seeking decisions that the rank comparison variables accounted for, across all symptoms, was 5.1% (range: 1-11%).

Likewise, across both studies and all symptoms, judgements of symptom severity and worry about symptoms were all significantly predicted by where participants thought their symptoms ranked in comparison to others. Rank variables accounted for, on average, across symptoms, 17.0% (range: 11-27%) and 16.3% (range: 9-28%) of the variance in symptom severity and worry judgements respectively. Given these significant associations and that these judgements have been shown to affect health help-seeking directly (e.g., Elliott et al., 2011; Martin et al., 2003), indirect effects between comparison of symptoms and help-seeking were also investigated. These showed that the associations between how participants believed their experience of symptoms ranked in comparison to that of others and whether or not they sought help for symptoms were fully mediated by judgements of symptom severity and worry about symptoms for all symptoms apart from colds, where the associations were partially mediated. In addition, further mediation analyses were undertaken in Study 1 that showed that, for depression, this association between rank of symptom occurrence and help-seeking was also fully mediated by participants' beliefs about whether they had depression and, for anxiety, was partially mediated by participants' beliefs about whether they had anxiety.

Symptom type and aspect: Comparison effects were seen consistently across both mental and physical health symptoms. The results suggest that people may compare the

frequency and/or the duration with which they experience symptoms to that of others when making judgements and decisions about their health.

Mechanisms underlying social comparison: Thus far, when summarising the comparison effects seen across the studies I have done so referring to the rank comparison variables only. This is because the results provided considerable evidence that when participants compared to others to make judgements and decisions about their symptoms they did so using rank-based strategies and not by comparing to the average of the comparison sample. The rank comparison variables were significant predictors of all of the outcomes measured across all six symptoms - a total of 38 models. The average amount of variance accounted for by these variables was 9.1% (range: 1-28%). In contrast, the distance from the average variables were only significant predictors in 12 out of the 38 models (32%) and accounted for, on average, only 1.2% (range: 0-1%) of explained variance in these 12 models. In 37 out of the 38 full models, the step 2a models (including the rank variables) either accounted for significantly more variance in the outcomes or there was very good evidence that the model fitted the data better than the step 2b models (including the distance from the average variables). Furthermore, in 31 out of the 38 step 2b models, rank was the strongest predictor of the outcomes, over and above participant's absolute experience of the symptom.

Accuracy of judgements and decisions: There was considerable variation in participants' beliefs about the average number of days in the last month (Study 1) or three months (Study 2) that people in the general population experience feeling depressed, anxious, tired or run down and muscle joint or back pain and the average length of time they experience headaches and colds for. This variation indicates that many participants had incorrect beliefs about the frequency and duration with which people experience these symptoms. Unsurprisingly given these findings, there was also considerable variation in participants' beliefs about where their symptom occurrence and duration ranked in

comparison to others among participants who experienced the same symptom occurrence or duration. This indicates that many participants misestimated their rank. This could affect the accuracy of health-related judgements and decisions if participants base these judgements and decisions on their beliefs about how their symptoms compare to those of others. Indeed, when help-seeking accuracy could be examined in Study 1, the findings showed that participants unlikely to have depression or anxiety were around four times more likely to seek help if they thought they experienced depression/anxiety symptoms more frequently than others. Participants likely to have clinical levels of depression or anxiety were two to three times more likely to not seek help if they thought they experienced symptoms less frequently than others.

Implications and Areas for Future Research

Symptom appraisal: The findings presented support previous research suggesting that people take into account the frequency with which they experience symptoms and the length of time that they experience them for when appraising their symptoms, i.e., making judgements about them such as how severe they are, how worried they are about them and whether they may be indicative of an underlying disorder (e.g., Arnault, 2009; Cauce et al., 2002; Elliott et al., 2011; Rickwood et al., 2005; Scott & Walter, 2010; Shaw et al., 2008). However, the results from both studies suggest that it is not the absolute experience of the symptom that matters most in this appraisal but where this experience is believed to rank in comparison to other people's experience of the symptom. In the vast majority of the regression models reported across both studies, the rank variables were the most important predictors of the judgement outcomes, over and above the absolute experience of the symptoms. When rank variables were added to the step 1 models including age, gender and the absolute experience, rank always either attenuated or removed the effect of the absolute experience. Furthermore, this pattern was consistent across all symptoms measured and both aspects of symptom experience. This suggests that social comparison

may be an integral part of symptom appraisal but, to my knowledge, this is not reflected in symptom appraisal models (see Whitaker, Scott, and Wardle, 2015, for a review of such models). The current studies may be the first to show the importance of social comparison in symptom appraisal but more research is needed to understand this process fully.

Help-seeking: Models of help-seeking generally concur that although help-seeking is influenced by many factors (Scott and Walter, 2010, suggest over 70), it is always preceded by some form of symptom appraisal which is likely to strongly influence help-seeking decisions (Cornally & McCarthy 2011; Henshaw & Freedman-Doan, 2009). Although many studies have shown that judgements such as those measured in this study do predict help-seeking behaviour, it has been acknowledged that little is known about how people actually interpret their symptoms and make these influential judgements about them (e.g., Henshaw & Freedman-Doan, 2009; Wills & Gibbons, 2009). The results from both studies provide a better understanding of the processes involved in this association between symptom appraisal and help-seeking. Where people believe their experience of symptoms ranks in comparison to other people's experience of the symptoms is linked directly to help-seeking decisions (the decision to seek help at all, the number of help-seeking behaviours to engage with and decisions to consult a professional and take medication) and indirectly through judgements of symptom severity, whether the symptom represents the presence of an underlying disorder and worry about symptoms. This suggests that individuals are evaluating their experience of symptoms by comparing to others and then using this evaluation to make help-seeking decisions directly and to make judgements about their symptoms that also influence help-seeking decisions.

The results not only provide a new understanding of how people evaluate their symptoms in order to make judgements and decisions about them, but also provide an explanation of how people can make inaccurate judgements and decisions. In the context of help-seeking, inaccurate decisions are made when people either fail to seek help when

they need to or do seek help when they do not need to. As discussed in the introduction, the majority of research in this area has focused on why the former occurs. Models such as the general model of total patient delay (Safer et al., 1979) suggest that many factors play a part in delaying seeking help when it is needed. Delay can occur at the symptom appraisal stage when people are evaluating whether they are ill, between making this decision and deciding whether professional care is needed, and between making the decision to seek care and actually going to the clinic or service. When these stages have been compared, appraisal delay has been found to contribute the most to the total delay time (e.g., Andersen et al., 1995; Ristvedt & Trinkaus, 2005). Making inaccurate judgements about not needing help at this stage can be costly and therefore it is important to understand why people may be doing so. The findings from the current studies show an association between beliefs about how symptoms compare to others and help-seeking accuracy such that participants were more likely to have not sought help when they needed to if they believed that they experienced symptoms less frequently than others. Similarly, participants were much more likely to seek help when they did not need to if they believed that they experienced symptoms more frequently than others. Therefore, inaccurate health help-seeking decisions may be made if an individual bases their decision on how they believe their health compares to that of others but their beliefs about others are not representative of the actual state of the world.

Henshaw and Freedman-Doan (2009) suggest that providing people with the information needed to make accurate symptom appraisals via public health messages and education campaigns could encourage more people that need help to seek it (and possibly discourage those that don't from help-seeking). The results reported here suggest that such campaigns should provide objective information about symptom experiences, such as telling people to seek help if they have experienced a symptom for more than a certain number of days, in order to discourage social comparison and the errors in judgement that

may occur because of it. Alternatively, information about the actual distribution of aspects of symptoms such as their occurrence or duration could be provided in order to encourage more accurate comparisons and hopefully more accurate decisions. Developing and testing the efficacy of such interventions could be a worthwhile area for future research.

Social comparison: As mentioned in the introduction, previous research on the use of social comparison in health-related decision-making has focused largely on the use of the lay referral structure for health advice with inaccuracy in help-seeking being explained as occurring due to receipt of inaccurate advice. The current studies suggest that social comparison plays a much more pivotal role, not only affecting help-seeking decisions but also influencing symptom evaluation and appraisal. These studies appear to be among the first to test whether people actually compare their symptoms to those of others when making related judgements and decisions and the first testing whether social comparison of symptoms can account for inaccurate health help-seeking behaviour. Motjabai (2008) showed an association between how people believed their mental distress compared to that of others (whether they thought they were more or less worried, nervous, or anxious than others) and help-seeking for mental health problems. The current studies support and extend these findings, showing associations between how people believe their experience of symptoms compares to that of others and a range of symptom-related judgements and help-seeking decisions. Mojtabai (2008) did not explicitly explore accuracy in help-seeking decisions as has been done here but his results suggested that people might have been more likely to seek help when it was not needed if they believed that they were more distressed than others. The current studies suggest that this could well be the case and that the same process may also cause people to not seek help when they need to because they believe that they are better off than others when in fact they are not.

The results also provide insights into the mechanisms underlying the comparison process. As explained in Chapter 1, it is largely assumed in the social comparison literature

that, when comparing to others, people do so through comparing to the average of the sample. The results from both studies suggest that, when making health-related judgements and decisions using social comparison, participants did so using rank-based strategies and not through comparing to the sample average.

Study Limitations

There were a number of limitations to the studies, the main one being the loss of the headache and cold symptom occurrence data due to participants not reading the questions properly. This was unfortunate as it would have been interesting to investigate which (if either) of the two symptom aspects, occurrence or duration, participants were more likely to compare to others when making judgements and decisions about their headaches and colds. It is possible that a better questionnaire design might have prevented this from happening. For example, if the headache and cold symptom occurrence questions had been placed on a separate page and more attention drawn to the fact participants were being asked about the number of headaches and colds they had had and not the number of days they had experienced these symptoms on as they had been for the tired/run down and muscle/joint/back pain questions.

A limitation of both studies was that the help-seeking measures were all self-reported, as it was not possible to obtain information from other sources such as medical records that could have verified help-seeking behaviours due to the large sample sizes needed and time limitations. Such sample sizes were needed as the majority of the comparison effects seen were small, as hypothesised. Similarly, the diagnostic tools used to measure whether participants had clinical levels of depression and anxiety so that their help-seeking accuracy could be investigated were also self-report measures. Ideally, this would have been determined using a diagnostic interview but again this was not possible due to the large number of participants. However, the PHQ-9 and GAD-7 are routinely used screening tools so as long as participants answered truthfully then this should not have

affected the results greatly. Comparison effects were still seen despite all of these limitations.

Conclusions

To conclude, two studies showed associations between where participants believed their experiences of six symptoms (feeling down, depressed or hopeless, nervous, anxious or on edge, tired or run down, muscle, joint or back pain, headaches, and cold symptoms) ranked in comparison to other people's experiences of these symptoms and judgements of symptom severity, worry about symptom and the presence of an underlying disorder (depression and anxiety only). Furthermore, direct associations between believed rank of symptom experience and help-seeking were also seen along with indirect associations through the above judgements. The results suggest that people compare to others and do so using rank-based strategies when they are appraising their symptoms and making help-seeking decisions. Help-seeking accuracy was also associated with believed rank of symptom experience – participants were more likely to seek help when it was not needed if they believed that they experienced symptoms more frequently than others and not to seek help when it was needed if they believed that they experienced symptoms less frequently than others. These comparison effects were small to medium in size.

The results have implications for models of symptom appraisal and help-seeking as they provide a new understanding of some of the mechanisms involved in these processes. They also provide a new understanding of how people can come to inaccurate help-seeking decisions and suggest ways in which inaccurate help-seeking may be targeted by information-based interventions. Finally, the studies provide evidence that people compare to others using rank-based strategies and show the existence of social comparison effects in a new area. Although it had previously been hypothesised that people may compare their experience of symptoms to those of others in order to make help-seeking decisions, this appears to have not been tested properly before.

Chapter 4: Sleep

Overview

Dysfunctional beliefs about sleep have been shown to exacerbate the kind of negative cognitive activity that leads to heightened arousal and emotional distress and ultimately sleep disturbance (Harvey, 2002). Such beliefs and their magnitude are typically identified and addressed in therapy for sleep disorders using the Dysfunctional Beliefs and Attitudes about Sleep scale (DBAS: Morin, 1993). Although this scale encompasses a wide range of sleep-related dysfunctional beliefs it is possible that other types of dysfunctional beliefs not included in the scale, such as how much one's sleep differs from an inaccurate perceived norm, may also play a role in this cycle. Therefore, the main aim of this study was to investigate whether there was any association between how people believe their sleep compares to that of others and sleep-related distress and worry. As in Chapter 3, this will be investigated by determining participants' beliefs about how their experience of sleep ranks in comparison to other people's sleep experiences and how much their experience differs from what they believe the average experience of others to be. Whether these beliefs influence help-seeking decisions for sleep problems will also be investigated.

Introduction

The American Psychiatric Association (2013) reports that around a third of the general population experience symptoms of insomnia (difficulty initiating sleep, maintaining sleep and waking up early and not being able to return to sleep), with 10-15% also experiencing associated daytime impairments (such as impairments to social, occupational, educational, academic or behavioural functioning), and 6-10% meeting diagnostic criteria for insomnia disorder (see Table 4.1 for full details). Consequences of insomnia include increased utilisation of health care services and consultations with health care professionals, increased use of medication (both prescription and over-the-counter)

and alcohol, increased accidents, increased absence from work, decreased productivity at work and poorer general health (Colten, Altevogt, & the Institute of Medicine Committee on Sleep Medicine and Research, 2006; Daley et al 2009). The estimated cost of insomnia alone to the US economy is around $63 billion a year (Kessler et al., 2011).

Furthermore, unhealthy sleep is associated with decreased health-related quality of life, increased morbidity and mortality, and the aetiology of many physical and mental health disorders. Much research from around the world has shown that people who routinely sleep for short or long durations (i.e. less than 7 hours a night or more than 9) are at increased risk of mortality (for a review see Grandner, Patel, Hale, & Moore, 2010), as are people who have been diagnosed with sleep disorders, especially sleep apnoea (Gooneratne et al., 2011). Habitual short and/or long sleep duration is associated with an increased risk of developing diabetes (Knutson, 2010; Zizi, et al., 2010), hypertension (Friedman, Bradley, Ruttanaumpawan, & Logan, 2010; Knutson, 2010) and weight problems (increased body mass index/obesity: Knutson, 2010; Nielsen, Danielsen, & Sorensen, 2010) as well as an increased risk of developing or dying from coronary heart disease, stroke and cardiovascular disease (Cappuccio, Cooper, D'Elia, Strazzullo, & Miller, 2011). High comorbidity between psychological and sleep disorders has been well documented (e.g., Morin & Ware, 1996). For example, around 40-50% of people with insomnia (the most prevalent sleep disorder) also have a psychological disorder (American Psychiatric Association, 2013). Insomnia is associated not only with the development of depression but also with relapse of episodes and resistance to treatment (Buysee et al., 2008; Ford & Kamerow, 1989; Ohayon, 2002; Okajima, Komada, Nomura, Nakashima, & Inoue, 2012). Insomnia is also associated with an increased risk of developing anxiety, stress, and problems related to substance abuse (Breslau, Roth, Rosenthal, & Andreski, 1996; Ford & Kamerow, 1989; Meerlo, Sgoifo, & Suchecki, 2008). Finally, using data from large-scale survey studies in the USA, France, and Japan, Léger et al. (2012) found that people with

insomnia had significantly lower health-related quality of life scores than good sleepers across countries.

Models of Insomnia

Many models of the development and maintenance of insomnia exist (e.g., the psychobiological inhibition model: Espie, Broomfield, MacMahon, Macphee, & Taylor, 2006; the stimulus control model: Bootzin, 1972; Harvey's 2002 cognitive model of insomnia; the neurobiological model of insomnia: Buysse, Germain, Hall, Monk, & Nofzinger, 2011; for a review see Buysee et al., 2011). These models are based on a number of distinct or integrated perspectives (e.g., biological, behavioural, cognitive, neuroscience). The diathesis-stress or '3P' model (Speilman, Caruso, & Glovinsky, 1987) and work by Ellis and colleagues (Ellis, Gehrman, Espie, Riemann, & Perlis, 2012; Ellis, Perlis, Neale, Espie, & Bastien, 2012) suggests that the transition from good sleeper to chronic insomnia begins with a trigger. This can be a perceived or actual stressor, such as a life event or series of events, which may be negative (e.g., a death, stress at work or illness) or positive in nature (e.g., planning a wedding). This can lead to sleep disturbance and the experience of acute insomnia (symptoms of insomnia for a period of between 3 days and 3 months - see Table 4.1 for a full definition). For most people (Espie, Perlis, et al. (2012) estimate 79%), normal sleep resumes after the initial trigger is resolved or managed, but for others acute insomnia develops into chronic insomnia. The model suggests that this may occur particularly to people who are predisposed to sleep difficulties - those who are female, older, more susceptible to hyperarousal and anxiety or who have a history of insomnia. Development of chronic insomnia is also likely to occur if people adopt behaviours (e.g., napping through the day, spending too much time in bed when not sleeping, drinking alcohol or caffeine) or cognitions (e.g., worrying about sleep and/or performance the next day) known to maintain or exacerbate sleep difficulties. As this study is concerned with the role that cognitions, particularly sleep-related beliefs, play in

influencing perceptions of sleep and sleep-related help-seeking behaviours, previous research on the role of cognition in sleep disturbance will be discussed briefly.

Harvey (2002) proposes a cognitive model of insomnia (and reviews the literature that supports it) whereby sleep disturbance is perpetuated by excessive, negatively toned cognitive activity. This could be, for example, uncontrollable worry and intrusive thoughts about such things as not getting enough sleep and the impact that this will have on health or functioning the next day. This cognitive activity then leads to autonomic arousal, emotional distress and anxiety, which is then likely to lead to an actual sleep deficit as sleep onset is unlikely under such conditions of high arousal (Bonnet & Arand, 1997). Furthermore, excessive cognitive activity during the day, such as ruminating about negative appraisals of the quality or duration of sleep obtained during the previous night, can also lead to physiological arousal and anxiety which is likely to impact on functioning that day.

The Role of Dysfunctional Beliefs in Sleep Disorder

Of the cognitive mechanisms thought to perpetuate unhealthy sleep, there has recently been a lot of interest in the role that sleep-related dysfunctional beliefs play in the maintenance of sleep disturbance. In Harvey's (2002) model, dysfunctional beliefs are shown to exacerbate the negative cognitive activity that leads to heightened arousal and emotional distress and subsequent sleep disturbance and/or deficits in daytime performance. As the model suggests, there are many studies that have shown how sleep-related dysfunctional beliefs contribute to the development (Jansson-Fröjmark & Linton, 2008), maintenance (Carney & Edinger, 2006; Edinger, Wohlgemuth, Radtke, Marsh, & Quillian, 2001b; Jansson-Fröjmark & Linton, 2007) and exacerbation (Morin, 1993; Morin, Blais, & Savard, 2002) of sleep disorders. As a result, the examination and challenge of sleep-related dysfunctional beliefs has been integrated into cognitive behavioural therapy for insomnia (CBT-I; a recommended standard and effective treatment for insomnia, Morgenthaler et al., 2006). There is also evidence to suggest that sleep-related

dysfunctional beliefs may be a factor in the sleep difficulties experienced by people with depression (Carney, Edinger, Manber, Garson, & Segal, 2007; Roecklein et al., 2013), seasonal affective disorder (Roecklein et al., 2013), fibromyalgia (Carney et al., 2007) and comorbid insomnia and mood disturbance (Carney et al., 2007).

The presence and strength of such beliefs are typically measured using the Dysfunctional Beliefs and Attitudes about Sleep scale (DBAS: Morin, 1993; Morin, Stone, Trinkle, Mercer, & Remsberg, 1993). The original version of the DBAS consists of 30 statements which respondents rate their level of agreement/disagreement with. There are also 16 (DBAS-16: Morin, Vallières, & Ivers, 2007) and 10 item (DBAS-10 or DBAS-SF: Espie, Inglis, Harvey, & Tessier, 2000) versions. The full DBAS contains statements relating to five aspects of sleep-related dysfunctional beliefs: misconceptions of the causes of insomnia (e.g., "I believe insomnia is essentially the result of a chemical imbalance"), misattributions or amplifications of the consequences of insomnia (e.g., "after a poor night's sleep, I know that it will interfere with my daily activities on the next day"), unrealistic sleep expectations (e.g., " I should sleep as well as my bed partner"), diminished perceptions of control (e.g., "when I sleep poorly on one night, I know it will disturb my sleep schedule for the whole week"), and faulty beliefs about sleep-promoting practices (e.g., "more time in bed ensures more sleep"). Depending on versions, respondents rate their level of agreement/disagreement with each of the statements on either a 100 mm analogue scale (strongly disagree to strongly agree) or a Likert scale from 0 (strongly disagree) to 10 (strongly agree). The DBAS is used to evaluate beliefs and identify cognitions that can be addressed in therapy and is used as an outcome measure of treatment.

Although this scale encompasses a wide range of sleep-related dysfunctional beliefs it is possible that other types of beliefs, such as how much one's sleep differs from the perceived norm, may also exacerbate the excessive negative cognitive activity that leads to emotional distress and ultimately disturbances in sleep and daytime functioning. For

example, Lack (2007) and Bruck, Dolan, and Lack (2015) describe norm-related inaccurate sleep beliefs held by participants in their studies. In both studies, the authors asked participants (both general population samples) to plot, on a chart, what they believed the normal sleep of both a healthy 18-year-old and a healthy 65-year-old to be. The chart had four blocks each underneath one another that were labelled 'awake', 'light sleep', 'deep sleep' and 'very deep sleep' and, on the top left, had the label 'night' and the top right the label 'morning'. In both studies, around 70% of participants plotted a U-shaped curve of unbroken sleep for both ages, with the average curve for the 18-year-old showing slightly deeper sleep than the average curve for the 65-year-old. Bruck et al. (2015) examined whether this response pattern differed across age groups and found that around 95% of younger adults and 75% of older adults believed that normal sleep for both 18 and 65-year-olds was gradual transition into and then out of deep sleep without any awakenings. The results indicate that many people hold the belief that a normal night's sleep involves continuous, deep sleep with no awakenings when, in fact, sleep occurs in cycles and it is quite normal to wake several times during the night (Ohayon, Carskadon, Guilleminault, & Vitiello, 2004). The authors conclude that it is likely that experiencing light sleep and/or awakenings during the night when holding the beliefs that most people sleep deeply and have unbroken sleep will cause anxiety, worry and possibly distress about sleep patterns, which may lead to sleep disturbance.

Possibly because of extensive use of the DBAS in this area there appears to have been little research apart from the above studies that has investigated people's perceptions of "normal" sleep. There has also been no research specifically looking at whether people compare aspects of their sleep to these perceived norms and the effect this may have on sleep-related worry or distress. Therefore, this study aims to do just that – to investigate whether people compare aspects of their sleep (such as how long they sleep for or the amount of time it takes them to get to sleep) to what they believe other

people experience in terms of these aspects and to see whether these comparisons are associated with how worried and distressed an individual is about their sleep.

Help-seeking

The second part of the study is concerned with investigating whether people's beliefs about how their sleep compares to that of others is associated with decisions regarding seeking help for sleep problems. This investigation is very similar to that conducted in the physical and mental health studies (reported in Chapter 3) in that both direct and indirect associations between compared sleep and help-seeking are explored. It is possible that, just as was found in these health studies, people may decide whether to seek help for sleep disturbances based on general perceptions and feelings they have about their sleep and/or how much they believe their sleep differs from perceived sleep norms.

Studies that have investigated determinants of sleep help-seeking behaviours have mainly focused on identifying symptoms and personal characteristics that predict help-seeking behaviours. These studies have identified that women, middle-aged and older adults, those with poorer general health and those with higher education, SES and self-efficacy to cope are the most likely to seek help for sleep difficulties (Ancoli-Isreal & Roth, 1999; Hohagen, et al., 1993; Hsu et al., 2013; Morin, LeBlanc, Daley, Gregoire, & Mérettee, 2006; Morin & Jarrin, 2013; Shochat, Umphress, Israel, & Ancoli-Israel, 1999). Morin et al. (2006) asked respondents in their survey who had previously sought help for sleep problems (n = 265) what had prompted them to do so, although they do not report if they were asked to select options from a list or write their answers freely. They found that the main determinants of help-seeking were daytime fatigue (48%), psychological distress (40%), and physical discomfort (22%) followed by suggestion by a significant other (14%), reduced work productivity (13%), suggestion by a health professional (11%), significant sleep loss (11%), self-referral (4%), and stress (3%).

As mentioned in previous chapters, improved knowledge about the determinants of help-seeing behaviour may help in understanding why people seek help when they do not need to and do not seek help when they do. Although there does not appear to be much research investigating this issue in the context of sleep help-seeking, there is some evidence that inappropriate help-seeking (either too much or too little) may be problematic. For example, using data from the National Health and Nutrition Examination Survey, 2005–2010, Chong, Fryar, and Gu (2013) report that people who sleep for nine hours a night use prescription sleep aids at around the same rate as those who sleep five hours or less (5.3% compared to 6.0% respectively) and that 4.1% of people who sleep for 8 hours take prescription sleep aids. It could be argued that these people sleep for so long because they are taking the medication but it is also possible that these figures indicate that good sleepers may be taking medication when they do not need to. If so, this phenomenon can be explained by people comparing themselves to inaccurate social norms. In the context of sleep, people may seek help for sleep difficulties when they do not need to if they believe that what they are experiencing is abnormal or may not seek help when they believe what they are experiencing is normal. For example, an individual may take, on average, 25 minutes to fall asleep but if they believe that the majority of others take 15 minutes or less to get to sleep they may think that their sleep onset latency is abnormal and seek help for this. On the other hand, someone who takes, on average, an hour to get to sleep may think this is perfectly normal if they believe that it takes the majority of people over an hour to get to sleep. It has been suggested that a sleep onset latency greater than 30 minutes indicates clinically significant severity (Ellis, Gehrman, et al., 2012; Lichstein, Durrence, Taylor, Bush, & Riedel, 2003) and so in this case, both individuals have come to incorrect judgements about the normality of their symptom because of the incorrect beliefs they hold about others. There is some anecdotal evidence that people may indeed compare to others to judge the severity of their sleep disturbance

and, consequently, whether or not they should seek help. In a qualitative study that assessed help-seeking in 26 individuals with chronic insomnia, Cheung, Bartlett, Armour, Glozier, and Saini (2014) report that, for participants who had a family history of sleep disturbance, "insomnia was embedded into the family culture, and appeared an acceptable way of life, which, therefore, did not warrant further investigation or medical help" (p. 112). This suggests that having people around you who also have insomnia almost normalises the experience leading to perceptions that help is not needed.

As mentioned earlier, indirect associations between social comparison and help-seeking will also be explored, using the same framework as in Chapter 3. As explained in that chapter, previous studies have shown that aspects of symptom appraisal, such as beliefs that symptoms represent the presence of a serious underlying disorder, judgements of symptom severity and worry or other emotional responses to symptoms, are also directly linked to seeking help for symptoms (e.g., Elliott et al., 2011; Martin et al., 2003). Findings from the mental and physical health studies supported this. It is hypothesised here that such appraisals – general perceptions and feelings about sleep – will also influence sleep help-seeking behaviours and that, as in the mental and physical health studies, social comparison will influence help-seeking directly and indirectly through these appraisals. The specific appraisals that are tested (based on previous research) are participants' beliefs about whether they have a sleep disorder, their perceptions of the quality of their sleep, how worried they are about their sleep and distress caused by sleeping problems.

There is some evidence to suggest that people do compare to others when making general judgements about their sleep such as whether they are good sleepers. Davis, Moore and Bruck (2006) asked 46 older adults living in community care homes to categorise themselves as either a good (n = 22) or poor (n = 24) sleeper and to complete the Pittsburg Sleep Quality Index (PSQI: Buysse, Reynolds, Monk, Berman, & Kupfer, 1989). This is a global measure of sleep quality that measures sleep onset latency, sleep duration,

habitual sleep efficiency, sleep disturbance, use of sleeping medication, subjective sleep quality, and daytime dysfunction. They found that all the self-categorised poor sleepers fell within the "poor" PSQI global score range but that over half of the self-categorised good sleepers fell within this range too. Based on conversations with the participants, the authors concluded that social comparison had contributed greatly to participants' self-categorisations. Self-categorised poor sleepers often made upward social comparisons or temporal comparisons to how they slept earlier on in their lives when no comparison group was available. Self-categorised good sleepers on the other hand mainly employed downward comparison strategies making them perceive that they were much better sleepers than they actually were.

In sum, based on the theory and findings outlined above, the following associations are tested:

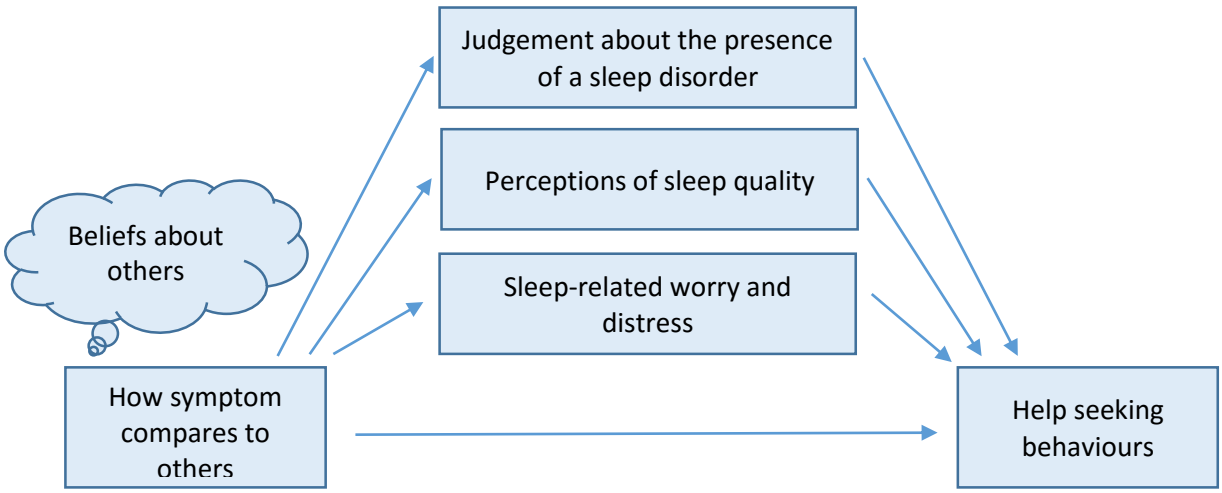

Figure 4.1. Hypothesised direct and indirect associations between social comparison and help-seeking.

Therefore, the full aims of the study are to:

1. Investigate whether judgements about sleep are in fact influenced by beliefs about how sleep compares to others and which specific aspects of sleep (such as sleep duration and the length of time taken to get to sleep) are compared,

2. Investigate whether worry and distress about sleep is associated with how people think their sleep compares to that of others,
3. Investigate whether people compare to others when making general judgements about their sleep such as whether they have a sleep problem and what they perceive the overall quality of their sleep to be,
4. Investigate whether people compare to others when making decisions about help-seeking for sleep problems either directly or indirectly through worry/distress about sleep and/or general sleep perceptions,
5. Investigate how people compare to others when making judgements and decisions about their sleep, namely whether they use rank or average-based comparisons as consistent with DbS and ALT respectively, and
6. Investigate whether there is any association between beliefs about how sleep compares to others and help-seeking accuracy.

Findings from this study may have implications for insomnia treatment, dysfunctional belief measurement and information and education campaigns. As previously mentioned, CBT-I is a standard treatment for insomnia and as its efficacy has been established (for a recent meta-analysis of RCTs see Okajima, Komada, & Inoue, 2011). Focus has recently turned to how treatment can be optimised for all insomnia patients. Sánchez-Ortuño and Edinger (2010) note that although the cognitive component of CBT targets dysfunctional beliefs, myths, and misconceptions about sleep, this tends to be done in quite a standardised way either by providing psychoeducation that targets specific, common misconceptions (Edinger, Wohlgemuth, Radtke, Marsh, & Quillian, 2001a) or through a standard cognitive restructuring protocol (Morin, Colecchi, Stone, Sood, & Brink, 1999). It is therefore possible that by using standardised methods some but not all dysfunctional beliefs and misconceptions are addressed. A more tailored approach could target all relevant misconceptions and may prove more beneficial to the patient. It is possible that,

as the DBAS is used extensively in CBT-I to identify dysfunctional beliefs, and as an outcome measure to measure change in these beliefs, other beliefs that could be causing the patient distress and perpetuating symptoms (such as variance from inaccurate perceived sleep norms) are not addressed in treatment. This study will investigate whether such beliefs may be a worthy target in CBT-I and whether they should be included in standard measures of dysfunctional beliefs.

As previously mentioned, dysfunctional beliefs can perpetuate sleep disturbance through causing distress about sleep but CBT-I interventions have shown that dysfunctional beliefs can be challenged effectively (e.g., Edinger et al., 2001b; Morin et al., 2002). Therefore, information and/or education campaigns that target challenging such beliefs may be effective in alleviating some of the distress that these beliefs may cause and, in turn, reduce sleep problems. If variance from inaccurately perceived sleep norms is found to cause distress about sleep, these norms could be targeted in campaigns that could explain what is 'normal' sleep in terms of hours of sleep per night, time taken to get to sleep and number of awakenings per night, etc. This may be of benefit not just to people with chronic insomnia but also to those with acute insomnia and good sleepers and may generally make people feel better about their sleep patterns and stop them from worrying so much about their sleep. Such information campaigns may also be useful to stop people from seeking help for sleep disorders when they do not need to and to encourage people who do need help to do so.

Table 4.1.
Diagnostic Criteria for Chronic and Acute Insomnia

	Acute Insomnia*	**Chronic Insomnia****
Trigger	1) Any life event or train of life events which results in a significant reduction in quality of life from the individual's ideal. 2) Distress at current situation.	1) Predominant complaint of dissatisfaction with sleep quantity or quality, associated with one (or more) of the symptoms below. 2) The sleep disturbance causes clinically significant distress or impairment in social, occupational, educational, academic, behavioural, or other important areas of functioning.
Symptoms	1. Difficulty initiating sleep. 2. Difficulty maintaining sleep, characterized by frequent awakenings or problems returning to sleep after awakenings. 3. Early-morning awakening with inability to return to sleep.	
Minimum frequency	3 or more nights per week	
Duration	3 days - 3 months	3 months
Course	Acute: 3-14 days. Transient: 2-4 weeks. Subchronic: 1-3 months.	Episodic: symptoms last at least 1 month but less than 3. Persistent: symptoms last 3 months or longer. Recurrent: two (or more) episodes within a year.
Additional criteria	1) The sleep difficulty occurs despite adequate opportunity for sleep. 2) The insomnia is not better explained by and does not occur exclusively during the course of another sleep-wake disorder (e.g., narcolepsy, a breathing-related sleep disorder, a circadian rhythm sleep-wake disorder, a parasomnia). 3) The insomnia is not attributable to the physiological effects of a substance (e.g., a drug of abuse, a medication). 4) Coexisting mental disorders and medical conditions do not adequately explain the predominant complaint of insomnia.	
Qualitative severity	Severity mild/moderate/severe as defined by the patient.	
Quantitative severity	+ 30 min sleep onset latency. + 30 min wake after sleep onset.	

*Note. *Definition taken from Ellis, Gehrman, et al. (2012), **definition taken from The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V, American Psychiatric Association, 2013).*

Method

As explained above, the main aims of the study were to investigate whether people compare their sleep to that of others when making sleep-related judgements and help-seeking decisions and to investigate whether there is any association between how participants believe their sleep compares to others and worry/distress about sleep. As in previous studies, how people compare to others and the extent to which their beliefs about others are correct are also investigated.

Participants

The 656 participants that completed this study were recruited through MTurk and had a mean age of 35.6 years (SD = 12.2, range: 18-75 years), were predominantly White (75%, Asian = 10%, Black = 7%, Hispanic/Latino = 4%, Native American or Alaskan Native = 1%, Other = 3%) and 57% were male. Participants were all resident in the U.S. and took part from 50 different states. Participants were mainly educated up to university (70%) or high school (28%) level (some high school = 1%, rather not say < 1%). Participants were mainly low to mid earners (< $14,999 = 21%, $15,000-$29,999 = 24%, $30,000-$49,999 = 25%, $50,000-$64,999 = 14%, >$65,000 = 14%, rather not say = 2%) and received $1.50 on completion of the study which took 10-15 minutes to complete.

Design and Procedure

The questionnaire that participants completed in this study was designed and hosted using Qualtrics. Participants were asked the following:

Sleep experiences: participants were asked the following about their experiences of five sleep aspects:

Sleep duration: "Over the last 90 days, how many hours of actual sleep did you get each night on average? (This may be different to the number of hours you spend in bed.)"

Sleep latency: "Over the last 90 days, how long (in minutes), on average, has it usually taken you to fall asleep each night?"

Sleep disturbance: "On how many of the last 90 nights have you had trouble sleeping because of waking up in the middle of the night or early morning?"

Daytime sleepiness: "On how many of the last 90 days have you had trouble staying awake during the day?"

Non-restorative sleep: "On how many of the last 90 days did you feel that your sleep was not refreshing and that you did not feel rested after sleeping?"

These aspects were chosen as they are commonly measured and have been found to be associated with general sleep perceptions and help-seeking behaviour in the literature (e.g., Harvey, Stinson, Whitaker, Moskovitz, & Virk, 2008; Ohayon, 2005; Morin et al., 2006). The wording of the first four items was taken from the Pittsburgh Sleep Quality Index (Buysee et al., 1989) which, as previously mentioned, is a measure of perceived sleep quality. The wording of the non-restorative sleep item was adapted from Ohayon (2005). As in the physical health study, a reporting period of 90 days was chosen to reduce the likelihood of floor effects.

Relative rank of sleep experience: Participants were asked directly where they thought their sleep experiences ranked amongst people their age in the general population through the questions below.

Sleep duration: "Out of 100 people, how many do you think have had more hours of actual sleep than you, on average, over the last 90 days?"

Sleep latency: "Out of 100 people, how many do you think have taken more time to get to sleep each night than you have, on average, over the last 90 days?"

Sleep disturbance: "Out of 100 people, how many do you think have had trouble sleeping because of waking up in the middle of the night or early morning on more nights than you over the last 90 days?"

Daytime sleepiness: "Out of 100 people, how many do you think have had trouble staying awake during the day on more days than you over the last 90 days?"

Non-restorative sleep: "Out of 100 people, how many do you think have felt that their sleep was not refreshing and that they did not feel rested after sleeping on more days than you over the last 90 days?"

This question format was used instead of the line placement procedure used in the mental and physical health studies reported in Chapter 3 as the procedure is much shorter in terms of explaining instructions to participants. Given that participants were asked about five different aspects, it was decided that a shorter question format would be more beneficial to the participants. Participants were asked to consider 'people' to be adults of about their age randomly selected from the U.S. population. Participants were asked to compare to people their age as studies have shown age-related changes in the sleep aspects included in this study (e.g., Ohayon et al., 2004).

Average sleep experience: Participants were asked what they thought the average experience of each sleep aspect was in people their age in the general population:

Sleep duration: "What do you think the average number of hours of sleep that people had each night over the last 90 nights was?"

Sleep latency: "What do you think the average time (in minutes) that it took people to fall asleep each night over the last 90 nights was?"

Sleep disturbance: "On average, on how many of the last 90 nights do you think people had trouble sleeping because of waking up in the middle of the night or early morning?"

Daytime sleepiness: "On average, on how many of the last 90 days do you think people had trouble staying awake during the day?"

Non-restorative sleep*:* "On average, on how many of the last 90 days do you think people felt that their sleep was not refreshing and that they did not feel rested after sleeping?"

The following outcomes were measured:

General sleep perceptions: Participants were asked the following general questions about their sleep:

Abnormality: "Do you think you have a sleep problem?" which participants answered using a slider scale that went from "definitely not" to "definitely yes".

Quality: "During the past 90 days, how would you rate your sleep quality overall?" which participants answered using a slider scale that went from "very bad" to "very good".

Worry: "During the past 90 days, how worried have you been about your sleep overall?" which participants answered using a slider scale that went from "not at all" to "a lot".

Distress: "During the past 90 days, how much distress have sleeping problems caused you?" which participants answered using a slider scale that went from "no distress" to "severe distress".

After each of these questions participants were asked which of the five sleep aspects ("the number of hours of sleep you get each night", "the length of time it takes you to fall asleep", "the number of times you wake up through the night or wake up early", "how often you have had trouble staying awake during the day", and "how often you wake up not feeling refreshed or rested after sleeping") they had considered the most when answering the question so that their absolute, rank and distance from the average values relating to that aspect could be used to predict the general judgements.

Specific sleep perceptions: In addition to the outcomes measuring general perceptions about sleep overall, outcome measures specific to each sleep aspect that participants were asked about were also measured. This was for two reasons: firstly, to test whether people compare any or all of these aspects when making judgements about their sleep – use of an outcome measure related to each aspect such as worry about the aspect allows for a test of social comparison effects using each aspect in isolation. Secondly, if

people are comparing sleep aspects, this methodology allows for a more comprehensive investigation of how people are comparing (i.e., whether they are using rank-based strategies or whether they compare to the average of the comparison sample). For brevity, the following specific questions were asked about two of the outcomes:

***Abnormality**:* "Based on the sleep you have had over the last 90 nights, do you think that you have a problem with any of the following?" participants were then presented with a list of the five aspects and gave their answer for each aspect on a slider scale from "definitely not" to "definitely yes".

***Worry**:* "Based on the sleep you have had over the last 90 nights, how worried are you about:" again, participants were presented with the same list of sleep aspects and gave their answer for each on a slider scale from "not at all" to "a lot".

Help-seeking behaviours: Participants were asked "which of these actions have you taken in the past 90 days to manage sleeping problems specifically? (Tick all that apply)" and given the following list:

- Looked for information,
- Discussed with other people,
- Contacted a charity or support group,
- Consulted a pharmacist,
- Consulted a nurse or doctor,
- Consulted a counselor or psychologist,
- Consulted any other professional (e.g., hypnotherapist, acupuncturist, homeopath, etc.),
- Taken over-the-counter medication,
- Taken prescribed medication,
- Undertaken cognitive behavioral therapy,
- Drank alcohol,
- Drank more caffeinated drinks during the day to combat tiredness after poor sleep,
- Used products (not including medication) that may aid sleep such as lavender spray, eye masks, ear plugs, soothing music, hypnotherapy CDs or self-help books,
- Taken specific actions that may aid sleep such as gone to bed early, taken a relaxing bath, meditated, read for a while, listened to music, relaxed, drank a warm, milky drink, stopped drinking caffeine after a certain time,
- None of the above.

If participants did not answer 'none of the above' they were also asked the following question "Please can you tell us which aspect of your sleep had the greatest influence on

your decision to take action to manage sleeping problems (please select one)" and selected their answer from a list of the five sleep aspects. These behaviours were taken in part from the physical and mental health studies reported in Chapter 3 and were adapted to include other known help-seeking behaviours specific to sleep problems such as those identified by Morin et al. (2006).

Insomnia screening: As in the mental health study in Chapter 3, participants were screened for the presence of insomnia (acute or chronic) in order to investigate the accuracy of the help-seeking decisions. The questions below, adapted from Ellis, Perlis, et al. (2012) and following the DSM-V diagnostic criteria for insomnia, as outlined in Table 4.1, were asked to determine the presence and degree (acute or chronic) of insomnia in participants. A standardised insomnia screening measure was not used in this study as one that a) distinguished between acute and chronic insomnia and b) met the current DSM-V diagnostic criteria could not be found. At the time of developing the study questionnaire, the two most commonly used sleep disorder screening tools were the aforementioned PSQI and the Insomnia Severity Index (ISI, Bastien, Vallières, & Morin, 2001). However, the PSQI lacks specificity for insomnia (it is a general measure of sleep disturbance) and the ISI lacks specificity for acute insomnia and is based on DSM-IV diagnostic criteria. The ability to identify cases of acute insomnia was deemed important for the current study as participants classed as having acute insomnia may not meet the threshold for chronic insomnia but are still experiencing significant sleep disturbance warranting help-seeking. Therefore, these participants may be classified as not having insomnia by the ISI but their help seeking behaviour may be appropriate as they have acute insomnia. The diagnostic criteria have also changed considerably between DSM-IV and DSM-V bringing the validity of the ISI into question (Ellis, Gehrman, et al., 2012). These issues have been discussed in the literature and, when measuring both acute and chronic insomnia, Ellis, Perlis, et al. (2012) devised their own screening tool based on asking questions that relate to the DSM-V

diagnostic criteria. Their tool is not a standardised measure and so its reliability and validity have not been tested meaning that there is no way of knowing how accurate it is as a screening tool. However, the questions closely follow the DSM-V diagnostic criteria (they are set out to resemble a diagnostic interview using these criteria) and allow for the identification of both acute and chronic insomnia which is why they were used in the current study. The questions are as follows:

1. "Have you ever had a problem with getting off to sleep, staying asleep or waking up too early?" Yes/No answer.

   [If yes to the first question]

2. "Is this an ongoing problem at the moment?" Yes/No answer.

   [If yes to the first and second questions]

3. "For how long has this been going on?" Participants saw three boxes to input the number of weeks/months/years. This allows for discrimination between acute and chronic insomnia.
4. "What is the nature of your sleep problem?" participants were asked to select all that apply from the following list: "Getting off to sleep", "staying asleep", "waking too early".
5. "Does your sleep problem result in significant distress or impairment in daytime functioning such as feeling fatigued or having low energy, feeling sleepy during the day, not being able to concentrate or remember things very well, disturbances in mood (e.g., irritability) or behaviour (e.g., aggression, impulsivity), impairment in functioning well at work or school, or impairment in social interaction?" Yes/No answer.
6. "Is your sleep problem present for 3 nights of the week or more?" Yes/No answer.
7. "Does your sleep problem occur despite adequate opportunity for you to sleep?" Yes/No answer.

8. “Is your sleep problem a side-effect of any substances you may be taking (e.g., medication or drugs?)” Yes/No answer.
9. “Do you have any of the following as diagnosed by a health professional?” participants selected all that applied from a list of disorders stated by the DSM-V to be comorbid with insomnia: Anxiety, arthritis, bipolar disorder, chronic obstructive pulmonary disease, coronary heart disease, depression, diabetes, fibromyalgia or any other chronic pain conditions.

All of the questions apart from the insomnia screening questions were ordered into two blocks of independent (sleep experience, rank, and average questions) and dependent variables (general sleep perceptions, specific sleep perceptions, and help-seeking questions). The order that these blocks were presented in was counterbalanced so that half the participants saw the independent variable block followed by the dependent variable block and half saw the dependent variable block followed by the independent variable block. This presentation order was random. Within each block the presentation order of the three sets of questions comprising each block was also randomised.

Following these two blocks, participants answered demographic questions (age, gender, level of education, ethnicity, income bracket and state in which they live), a question asking if they had a sleep disorder as diagnosed by a medical professional (and if so which one) and then completed the insomnia screening questionnaire.

Results

Initial data screening showed that the absolute and distance from the average duration and latency variables contained a number of outlying cases; see Figure 4.2 below. The duration and latency rank variables were not affected as participants’ answers were controlled through use of a slider scale. Similarly, as participants answered on how many of the last 90 days or nights they had experienced the three other sleep aspects and their

answers were limited to numbers less than or equal to 90, there were no outlying cases in any of the independent variables measuring these aspects.

On inspection of the data it was decided that, for the duration variables, participants with a sleep duration and/or distance from the average duration value that was greater or less than ±3 standard deviations from the mean would be excluded from the analysis. This equated to sleep duration values less than 1.71 hours and greater than 11.73 hours and distance from the average duration values less than -4.72 hours and greater than 5.10 hours. This resulted in eight exclusions: P31 (sleep duration = 1 hour), P13, P349, P480, P635 (sleep duration all = 12 hours), P460 (sleep duration = 23 hours; distance from the average = -16 hours), P108 (sleep duration = 24 hours) and P493 (distance from the average = 6 hours). Applying the same exclusion criteria to the latency data would have resulted in the exclusion of 31 participants and therefore it was decided to just remove the four most extreme cases. These were: P572 (sleep latency = 300 minutes), P495 (sleep latency = 480 minutes; distance from the average = -435 minutes), P108 (already excluded because of their sleep duration; distance from the average = 380 minutes) and P44 (distance from the average = 403 minutes). Therefore, a total of 11 participants were excluded from all analyses.

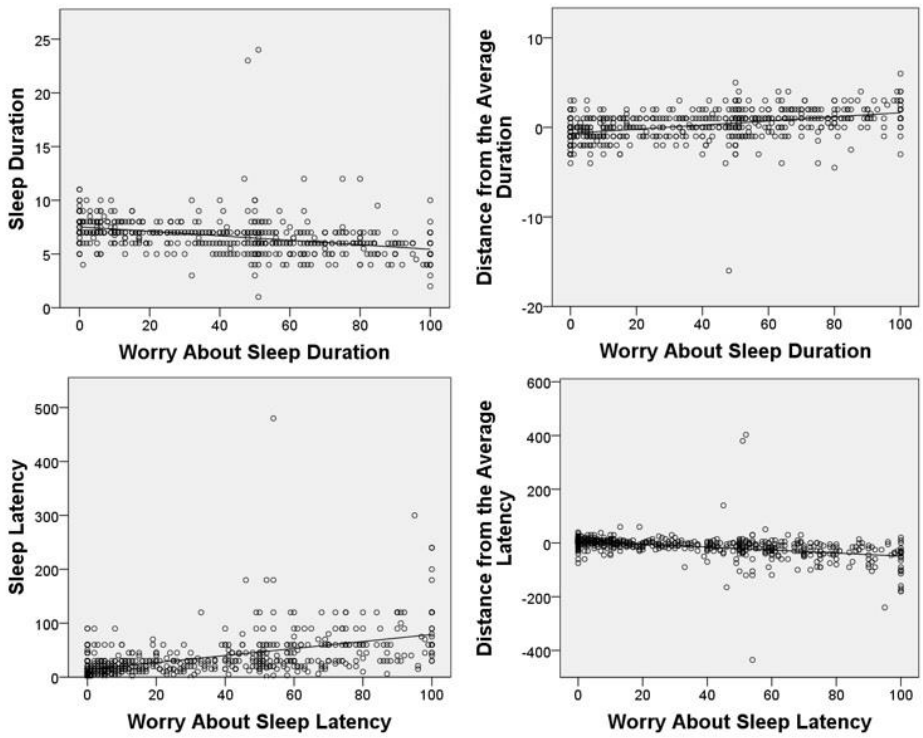

Figure 4.2. Scatterplots of participants' absolute (sleep duration and latency) and distance from the average values against their worry scores for each sleep aspect to depict outlying cases.

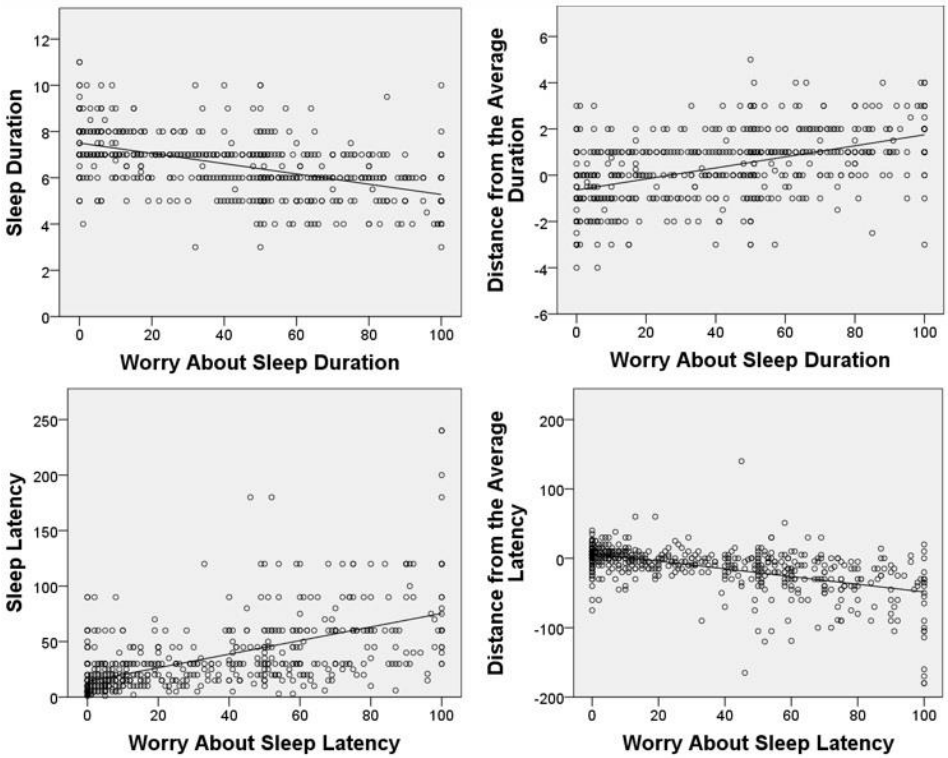

Figure 4.3. Scatterplots of participants' absolute (sleep duration and latency) and distance from the average values against their worry scores for each sleep aspect after participant exclusion.

Figure 4.3 above shows the same scatterplots as Figure 4.2 after these participants were excluded. As suggested by Figure 4.3 and as shown in Table 4.2, correlations between the absolute and distance from the average variables and their corresponding dependent variables became stronger after the participants were excluded. Even though they were unaffected by outliers, the duration and latency rank variables are also included in Table 4.2 to show the differences in the correlations between these variables and the outcomes when the participants were excluded.

Table 4.2.
Correlations between Participants' Sleep Duration, Sleep Latency, the Distance of their Sleep Duration and Sleep Latency from their Believed Population Average Sleep Duration and Latency, the Rank of their Sleep Duration and Sleep Latency, and Outcome Variables Relating to these Measures, Before and After Participants were Excluded.

	All Participants		After Exclusion	
	Worry About Duration	Problem with Duration	Worry About Duration	Problem with Duration
Sleep Duration	-.400***	-.420***	-.548***	-.560***
Distance from the Average Duration	-.445***	-.454***	-.518***	-.524***
Rank of Duration	-.412***	-.423***	-.425***	-.434***
N	656	656	645	645
	Worry About Latency	Problem with Latency	Worry About Latency	Problem with Latency
Sleep Latency	.539***	.572***	.607***	.634***
Distance from the Average Latency	.421***	.462***	.576***	.608***
Rank of Latency	.472***	.517***	.460***	.517***
N	656	656	645	645

*Note. ***p < .001*

This results section addresses each of the aims outlined in the introduction on pages 151 and 152 in turn, although the use of rank or average-based comparison strategies (aim 5) is discussed throughout. Correlations conducted prior to the main analyses showed generally that participants' absolute experiences of the aspects correlated highly with the corresponding distance from the average variables (range: r = .721 to .908). However, tolerance values obtained through running multiple linear regression analyses confirmed that collinearity is not problematic in any of the models including absolute and distance from the average variables (range of tolerance values for absolute variables: .214 to .165, distance from the average variables: .218 to .469). To recap, the main independent variables measured were as follows:

Absolute experience of five sleep aspects: This was participants' actual experiences of the five sleep aspects:

1. The average number of hours sleep participants had over the last 90 days (sleep duration: *M* = 6.69, *SD* = 1.29),
2. The average length of time in minutes it took participants to fall asleep over the last 90 days (sleep latency: *M* = 32.31, *SD* = 31.94),
3. The number of nights out of the last 90 that participants had trouble sleeping because of waking up in the middle of the night or early morning (sleep disturbance: *M* = 29.66, *SD* = 30.28),
4. The number of nights out of the last 90 that participants had trouble staying awake during the day (daytime sleepiness: *M* = 19.45, *SD* = 25.00), and
5. The number of nights out of the last 90 that participants had felt that their sleep was not refreshing and that they did not feel rested after sleeping (non-restorative sleep: *M* = 36.60, *SD* = 31.49).

Relative rank of sleep experience: This was the number of people out of 100 who:

1. Had more hours of actual sleep than the participant (duration rank: *M* = 49.36, *SD* = 24.99),
2. Took more time to get to sleep each night (latency rank: *M* = 51.94, *SD* = 27.02),
3. Had trouble sleeping because of waking up in the middle of the night or early morning on more nights than them (disturbance rank: *M* = 49.59, *SD* = 25.81),
4. Had trouble staying awake during the day on more days than them (daytime sleepiness rank: *M* = 49.06, *SD* = 27.72), and
5. Felt that their sleep was not refreshing and that they did not feel rested after sleeping on more days than them (non-restorative sleep rank: *M* = 50.97, *SD* = 25.65) over the last 90 days.

Distance from the average sleep experience: Participants' answers to the questions asking what they though the average experience of the five sleep aspects in people from the general population about their age was were deducted from their answers to the relevant questions asking about their absolute experience of these aspects. This gave a direct measure of how much participants' experiences of the aspects differed from what they believed the average experience of others to be (duration distance from the average: *M* = -.024, *SD* = 1.46, latency distance from the average: *M* = 9.42, *SD* = 30.48, disturbance distance from the average: *M* = 6.11, *SD* = 31.06, daytime sleepiness distance from the average: *M* = -3.61, *SD* = 25.92, non-restorative sleep distance from the average: *M* =2.44, *SD* = 32.52).

Aim 1

The first aim of the study was to investigate whether judgements about sleep are influenced by beliefs about how sleep compares to others and which specific aspects of sleep are compared. To recap, the judgements used to investigate this were 1) whether the participant thought they had a problem with their experience of the aspect, measured on a 0 to 100 scale from "definitely not" to "definitely yes" and 2) how worried the participant was about their experience of the aspect, measured on a 0 to 100 scale from "not at all" to "a lot". Table 4.3 below shows the means and their associated standard deviations of participants' responses.

Table 4.3.
Descriptive Statistics of Participants' Responses to Questions Asking Whether They Thought They Had a Problem With Any Sleep Aspect and How Worried They Were About Their Experience of Each Aspect

	Problem		Worry	
Aspect	Mean	Std. Deviation	Mean	Std. Deviation
Sleep Duration	45.34	34.37	36.69	31.75
Sleep Latency	37.78	34.79	29.63	31.67
Sleep Disturbance	42.29	34.09	35.78	32.79
Daytime Sleepiness	30.60	31.33	28.32	31.71
Non-restorative Sleep	49.11	34.29	44.80	34.61

Note. *N* = 645.

Table 4.4.

Multiple Linear Regression Analyses Predicting Worry about Sleep Duration Ratings and Whether Participants think they have a Problem with their Sleep Duration from Age, Gender, Sleep Duration, Rank of Sleep Duration, and Distance from the Believed Average Sleep Duration of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Worry					Sleep Problem				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.32***				**<.001**	.32***				**<.001**
Constant		138.23 (124.79-152.02)	6.25		.001		147.04 (132.75-161.60)	7.20		.001
Age		**-0.31 (-0.47--0.15)**	**0.08**	**-.12**	**.001**		-0.10 (-0.29-0.09)	0.09	-.04	.278
Gender		1.17 (-3.77-6.24)	2.21	.02	.583		4.20 (0.05-8.40)	2.22	.06	.057
Sleep Duration		**-13.60 (-15.15--12.10)**	**0.85**	**-.55**	**.001**		**-14.95 (-16.92--13.11)**	**1.00**	**-.56**	**.001**
Step 2a	.02***				**<.001**	.03***				**<.001**
Constant		132.62 (119.53-145.95)	6.30		.001		140.56 (126.02-155.14)	7.28		.001
Age		**-0.30 (-0.44--0.14)**	**0.08**	**-.12**	**.001**		-0.08 (-0.25-0.09)	0.09	-.03	.355
Gender		1.85 (-2.88-6.64)	2.18	.03	.388		**4.98 (0.72-9.03)**	**2.17**	**.07**	**.023**
Sleep Duration		**-11.22 (-13.20--9.44)**	**1.02**	**-.46**	**.001**		**-12.20 (-14.40--10.10)**	**1.15**	**-.46**	**.001**
Rank		**-0.23 (-0.33--0.12)**	**0.05**	**-.18**	**.001**		**-0.26 (-0.37--0.16)**	**0.06**	**-.19**	**.001**
Step 2b	.01**				**<.001**	.01**				**<.001**
Constant		109.28 (89.12-130.28)	10.30		.001		116.47 (92.67-139.46)	10.76		.001
Age		**-0.31 (-0.47--0.13)**	**0.08**	**-.12**	**.001**		-0.10 (-0.27-0.08)	0.09	-.03	.278
Gender		1.64 (-2.64-6.22)	2.13	.03	.447		**4.70 (0.18-9.56)**	**2.28**	**.07**	**.048**
Sleep Duration		**-9.48 (-12.28--6.93)**	**1.47**	**-.38**	**.001**		**-10.59 (-13.55--7.79)**	**1.57**	**-.40**	**.001**
Distance from the Average		**-4.35 (-6.99--1.79)**	**1.24**	**-.20**	**.002**		**-4.60 (-6.92--1.96)**	**1.32**	**-.20**	**.003**

*Note. N = 645, CI = confidence interval. Significant models and predictors are highlighted in bold. **p<.01, ***p<.001*

Table 4.5.

Multiple Linear Regression Analyses Predicting Worry about Sleep Latency Ratings and Whether Participants think they have a Problem with their Sleep Latency from Age, Gender, Sleep Latency, Rank of Sleep Latency, and Distance from the Believed Average Sleep Latency of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Worry					Sleep Problem				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.40***				**<.001**	.42***				**<.001**
Constant		27.17 (21.42-32.29)	3.23		.001		30.64 (23.94-37.16)	3.74		.001
Age		**-0.43 (-0.58--0.28)**	**0.07**	**-.17**	**.001**		**-0.39 (-0.56--0.21)**	**0.08**	**-.14**	**.001**
Gender		-2.92 (-6.98-1.21)	1.99	-.05	.144		-1.94 (-6.19-2.51)	2.22	-.03	.397
Sleep Latency		**0.59 (0.52-0.69)**	**0.04**	**.60**	**.001**		**0.68 (0.59-0.79)**	**0.05**	**.63**	**.001**
Step 2a	.05***				**<.001**	.07***				**<.001**
Constant		14.25 (7.90-20.16)	3.35		.001		13.48 (6.86-19.72)	3.77		.002
Age		**-0.40 (-0.55--0.25)**	**0.07**	**-.16**	**.001**		**-0.35 (-0.51--0.18)**	**0.08**	**-.12**	**.001**
Gender		-2.06 (-6.00-2.09)	1.91	-.03	.273		-0.81 (-4.82-3.19)	2.09	-.01	.708
Sleep Latency		**0.49 (0.41-0.59)**	**0.04**	**.49**	**.001**		**0.54 (0.46-0.65)**	**0.05**	**.50**	**.001**
Rank		**0.28 (0.21-0.36)**	**0.04**	**.24**	**.001**		**0.38 (0.29-0.46)**	**0.04**	**.29**	**.001**
Step 2b	.01**				**<.001**	.01***				**<.001**
Constant		30.95 (24.85-37.89)	3.45		.001		35.59 (26.62-45.65)	4.36		.001
Age		**-0.44 (-0.57--0.31)**	**0.07**	**-.17**	**.001**		**-0.41 (-0.57--0.24)**	**0.08**	**-.14**	**.001**
Gender		-2.61 (-6.53-1.43)	2.01	-.04	.209		-1.53 (-5.61-2.68)	2.15	-.02	.494
Sleep Latency		**0.43 (0.23-0.56)**	**0.08**	**.43**	**.001**		**0.46 (0.22-0.62)**	**0.11**	**.42**	**.001**
Distance from the Average		**0.20 (0.08-0.39)**	**0.08**	**.19**	**.015**		**0.26 (0.09-0.57)**	**0.11**	**.23**	**.015**

*Note. N = 645, CI = confidence interval. Significant models and predictors are highlighted in bold. **p<.01, ***p<.001*

Table 4.6.

Multiple Linear Regression Analyses Predicting Worry about the Frequency of Sleep Disturbance Ratings and Whether Participants think they have a Problem with the Frequency of their Sleep Disturbance from Age, Gender, Frequency of Sleep Disturbance, Rank of the Frequency of Sleep Disturbance, and Distance from the Believed Average of the Frequency of Sleep Disturbance of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Worry					Sleep Problem				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.39***				**<.001**	.52***				**<.001**
Constant		22.68 (16.08-29.41)	3.29		.001		20.63(14.69-27.41)	3.02		.001
Age		**-0.19 (-0.36--0.01)**	**0.09**	**-.07**	**.034**		-0.08 (-0.26-0.08)	0.09	-.03	.325
Gender		-1.33 (-5.38-2.71)	2.11	-.02	.545		0.76 (-3.27-4.63)	1.97	.01	.710
Sleep Disturbance		**0.69 (0.62-0.750**	**0.03**	**.64**	**.001**		**0.82 (0.76-0.87)**	**0.03**	**.73**	**.001**
Step 2a	.02***				**<.001**	.01***				**<.001**
Constant		15.53 (8.53-22.34)	3.51		.001		13.64 (7.33-20.66)	3.12		.001
Age		**-0.19 (-0.36--0.01)**	**0.09**	**-.07**	**.024**		-0.08 (-0.26-0.08)	0.09	-.03	.319
Gender		-0.93 (-5.01-3.11)	2.10	-.01	.671		1.15 (-2.74-4.90)	1.94	.02	.579
Sleep Disturbance		**0.65 (0.58-0.72)**	**0.04**	**.60**	**.001**		**0.78 (0.72-0.84)**	**0.03**	**.70**	**.001**
Rank		**0.16 (0.09-0.24)**	**0.04**	**.13**	**.001**		**0.16 (0.10-0.23)**	**0.04**	**.12**	**.001**
Step 2b	.00				**<.001**	.00				**<.001**
Constant		22.93 (15.94-30.46)	3.54		.001		21.49 (16.02-27.28)	2.98		.001
Age		**-0.19 (-0.37--0.02)**	**0.09**	**-.07**	**.048**		-0.08 (-0.23-0.08)	0.08	-.03	.318
Gender		-1.28 (-5.76-3.14)	2.09	-.02	.528		0.93 (-2.49-4.58)	1.91	.01	.639
Sleep Disturbance		**0.67 (0.54-0.80)**	**0.07**	**.62**	**.001**		**0.77 (0.65-0.87)**	**0.06**	**.68**	**.001**
Distance from the Average		0.02 (-0.10-0.14)	0.06	.02	.776		0.06 (-0.04-0.17)	0.06	.05	.292

*Note. N = 645, CI = confidence interval. Significant models and predictors are highlighted in bold. ***p<.001*

Table 4.7.

Multiple Linear Regression Analyses Predicting Worry about the Frequency of Daytime Sleepiness Ratings and Whether Participants think they have a Problem with the Frequency of their Daytime Sleepiness from Age, Gender, Frequency of Daytime Sleepiness, Rank of the Frequency of Daytime Sleepiness, and Distance from the Believed Average of the Frequency of Daytime Sleepiness of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Worry					Sleep Problem				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.52***				**<.001**	.54***				**<.001**
Constant		20.66 (15.23-25.89)	2.70		.001		20.91(15.49-26.40)	2.84		.001
Age		**-0.28 (-0.41--0.13)**	**0.07**	**-.11**	**.001**		**-0.23 (-0.36--0.08)**	**0.07**	**-.09**	**.003**
Gender		-0.02 (-3.49-3.68)	1.77	.00	.993		-0.03 (-3.41-3.43)	1.72	.00	.988
Daytime Sleepiness		**0.91 (0.83-0.98)**	**0.04**	**.71**	**.001**		**0.92 (0.84-0.99)**	**0.04**	**.73**	**.001**
Step 2a	.02***				**<.001**	.01***				**<.001**
Constant		14.63 (9.23-19.98)	2.65		.001		15.20 (9.48-21.16)	2.77		.001
Age		**-0.29 (-0.42--0.15)**	**0.07**	**-.11**	**.001**		**-0.24 (-0.38--0.09)**	**0.07**	**-.09**	**.001**
Gender		0.27 (-3.25-3.90)	1.74	.00	.869		0.24 (-3.12-3.68)	1.71	.00	.892
Daytime Sleepiness		**0.87 (0.79-0.96)**	**0.05**	**.69**	**.001**		**0.88 (0.80-0.96)**	**0.04**	**.71**	**.001**
Rank		**0.14 (0.09-0.20)**	**0.03**	**.13**	**.001**		**0.14 (0.08-0.19)**	**0.03**	**.12**	**.001**
Step 2b	.00				**<.001**	.00				**<.001**
Constant		20.93 (15.38-26.73)	2.93		.001		20.99 (15.88-25.92)	2.96		.001
Age		**-0.28 (-0.42--0.14)**	**0.07**	**-.11**	**.002**		**-0.23 (-0.37--0.08)**	**0.07**	**-.09**	**.004**
Gender		0.00 (-3.39-3.69)	1.82	.00	.999		-0.03 (-3.45-3.61)	1.73	.00	.984
Daytime Sleepiness		**0.89 (0.75-1.04)**	**0.07**	**.71**	**.001**		**0.91 (0.80-1.03)**	**0.06**	**.73**	**.001**
Distance from the Average		0.01 (-0.10-0.13)	0.06	.01	.794		0.00 (-0.10-0.10)	0.05	.00	.932

*Note. N = 645, CI = confidence interval. Significant models and predictors are highlighted in bold. ***p<.001*

Table 4.8.

Multiple Linear Regression Analyses Predicting Worry about the Frequency of Non-Restorative Sleep Ratings and Whether Participants think they have a Problem with the Frequency of Non-Restorative Sleep from Age, Gender, Frequency of Non-Restorative Sleep, Rank of the Frequency of Non-Restorative Sleep, and Distance from the Believed Average of the Frequency of Non-Restorative Sleep of People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Worry					Sleep Problem				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.53***				**<.001**	.56***				**<.001**
Constant		26.74 (20.63-33.63)	3.15		.001		31.48 (25.39-38.01)	3.03		.001
Age		**-0.30 (-0.44--0.18)**	**0.07**	**-.11**	**.001**		**-0.33 (-0.46--0.20)**	**0.07**	**-.12**	**.001**
Gender		0.14 (-3.84-3.96)	1.98	.00	.948		-0.42 (-3.73-3.03)	1.79	-.01	.817
Non-Restorative Sleep		**0.79 (0.73-0.84)**	**0.03**	**.71**	**.001**		**0.81 (0.75-0.86)**	**0.03**	**.74**	**.001**
Step 2a	.01**				**<.001**	.01***				
Constant		21.07 (14.25-28.01)	3.45		.001		25.78 (19.34-32.72)	3.26		.001
Age		**-0.30 (-0.44--0.18)**	**0.07**	**-.11**	**.001**		**-0.32 (-0.46--0.20)**	**0.07**	**-.12**	**.001**
Gender		0.31 (-3.56-4.32)	1.96	.00	.877		-0.25 (-3.58-3.10)	1.77	.00	.899
Non-Restorative Sleep		**0.76 (0.69-0.82)**	**0.03**	**.69**	**.001**		**0.78 (0.72-0.83)**	**0.03**	**.71**	**.001**
Rank		**0.13 (0.05-0.20)**	**0.04**	**.10**	**.002**		**0.13 (0.06-0.20)**	**0.03**	**.10**	**.001**
Step 2b	.00				**<.001**	.00				
Constant		27.19 (20.67-33.55)	3.22		.001		31.75 (25.38-38.23)	3.36		.001
Age		**-0.30 (-0.44--0.16)**	**0.07**	**-.11**	**.001**		**-0.33 (-0.46--0.18)**	**0.07**	**-.12**	**.001**
Gender		0.23 (-3.31-3.85)	1.95	.00	.908		-0.36 (-4.29-3.39)	1.91	-.01	.826
Non-Restorative Sleep		**0.77 (0.68-0.86)**	**0.04**	**.70**	**.001**		**0.80 (0.70-0.89)**	**0.05**	**.73**	**.001**
Distance from the Average		0.02 (-0.08-0.11)	0.05	.02	.674		0.01 (-0.07-0.10)	0.04	.01	.784

*Note. N = 645, CI = confidence interval. Significant models and predictors are highlighted in bold. **p<.01, ***p<.001*

Multiple linear regression was used to investigate whether the rank and distance from the average variables for each aspect predicted the corresponding judgements and these analyses are reported in Tables 4.4 to 4.8 above. These tables show the same pattern of results across the two judgements for each aspect in terms of the significance of the absolute, rank, and distance from the average variables. For all five aspects of sleep, participants' absolute experience of the aspect[6] and where they believed their experiences ranked in comparison to other people's experiences of the aspects were significant predictors of both worry about their experience of the aspect and whether they thought their experience was problematic/abnormal. How much participants' experiences of aspects differed from what they believed the average experience of the aspect in people their age from the general population to be significantly predicted worry and abnormality judgements for sleep duration and latency only. The sleep disturbance, daytime sleepiness and non-restorative sleep distance from the average variables did not predict worry or abnormality judgements about these aspects.

There is also a very similar pattern across all aspects and judgements in the strength of the predictors in the models. Participants' absolute experience of the sleep aspects is always the most important predictor of worry and abnormality judgements, in all of the model steps. The results reported in Chapter 3 showed a general trend across both studies whereby the effect of the absolute experience (symptom occurrence or duration) was considerably attenuated or, in some cases, removed when rank variables were entered into the models at step 2a. In the majority of cases, this lead to the rank variables being the most important predictors of the outcomes measured. This is not seen here. There is some attenuation of the effect of the absolute when rank is added but it is nowhere near the

[6] i.e., the average number of hours sleep they got and time they took to get to sleep over the last 90 days and the number of days in the last 90 that they had trouble sleeping because of waking up throughout the night, trouble staying awake during the day and felt that their sleep was not refreshing.

degree seen in the previous studies. Furthermore, in all of the models reported in the tables above, the absolute variables are much stronger predictors of the outcome than the rank variables (the absolute beta weights were significantly greater than the rank beta weights in all cases). This is also the case with the distance from the average variables; in the models where they are significant predictors of worry and abnormality judgements, they are significantly weaker predictors than the absolute variables.

The step 1 models including age, gender and participants' absolute experience of the sleep aspect account for, on average, 43% of the variance in aspect specific worry judgements and 47% of the variance in aspect specific abnormality judgements. The addition of the rank variables at step 2a always results in a significant increase in the amount of explained variance but the amount added is small; 2-3% on average for both judgements. When the distance from the average variables are entered into both duration and latency models at step 2b, the amount of explained variance increases significantly but by only 1% in all four models. Comparison of the predictive ability of the step 2a and 2b models showed that the step 2a models accounted for significantly more variance in latency, sleep disturbance and daytime sleepiness worry and abnormality judgements and non-restorative sleep abnormality judgements than the step 2b models[7]. There was no significant difference in the amount of variance in sleep duration worry and abnormality judgements or non-restorative sleep worry judgements accounted for by the step 2a and step 2b models, duration worry: $t(642) = 1.05$, $p = .148$, duration abnormality: $t(642) = 1.23$, $p = .109$, non-restorative worry: $t(642) = 1.11$, $p = .134$.

In sum, the results provide evidence that participants compared their experience of all of the five aspects measured to other people's experiences of these aspects when

[7] Latency worry: t(642) = 3.09, p = .001, latency abnormality: t(642) = 4.00, p < .001, disturbance worry: t(642) = 2.19, p = .014, disturbance abnormality: t(642) = 2.08, p = .019, sleepiness worry: t(642) = 2.22, p = .013, sleepiness abnormality: t(642) = 2.36, p = .009, non-restorative abnormality: t(642) = 1.97, p = .025.

making judgements about how worried they were about their experience of each aspect and whether they thought their experience of the aspect was problematic. The results suggest that, for the majority of aspects (sleep latency, disturbance, daytime sleepiness and non-restorative sleep), participants compared using rank-based strategies; the results were inconclusive as to how participants compare to others when making judgements about their sleep duration. Comparison effects were small but significant, and the most important predictor of aspect specific judgements was participants' absolute experience of the aspect.

Aim 2

As the above results have shown that participants' beliefs about how aspects of their sleep compare to others influence their judgements about all five aspects measured, it is possible to explore whether overall worry and distress relating to sleep is associated with how people think their sleep compares to that of others using data from all five aspects. To recap, participants were asked how worried they had been about their sleep overall during the last 90 days (measured on a 0 to 100 scale from "not at all" to "a lot", M = 38.04, SD = 29.46) and how much distress sleeping problems had caused them during the last 90 days (measured on a 0 to 100 scale from "no distress" to "severe distress", M = 28.55, SD = 26.71). Participants' answers to these questions were predicted from the absolute, rank, and distance from the average values of the aspect each participant stated that they had referred to the most when making these two judgements. In order to do this, new absolute, rank, and distance from the average variables were constructed using the following procedure:

1. Participants' absolute, rank, and distance from the average values for all five aspects were standardised (transformed into z scores) as the aspects were all measured using different scales (hours, minutes, days). This only affected the absolute and distance

from the average variables as rank was measured on a 0 to 100 scale for all aspects but, for consistency, all variables were standardised.

2. Absolute, rank, and distance from the average values for the sleep duration aspect were inversed (i.e., multiplied by -1) so that, for all aspects, low absolute, rank, and distance from the average values indicated good sleep (e.g., long duration, short latency, few awakenings, etc.) and high values indicated bad sleep (e.g., short duration, long latency, numerous awakenings, etc.).
3. For the following general outcomes: whether participants thought they had a sleep problem, sleep quality, worry about sleep, distress caused by sleep problems and help-seeking, outcome specific absolute, rank and distance from the average variables were compiled. This was done using the transformed values from the aspect that each participant stated that they referred to the most when answering the question about that specific outcome. For example, one participant said that they had predominantly based their answer to the sleep quality question on the number of days that they felt refreshed after sleeping but had mainly based their answer to the worry about sleep question on the average number of hours of sleep they tend to get. In this case, their non-restorative sleep absolute, rank, and distance from the average answers would be used to predict their quality rating and their duration answers would be used to predict their worry rating.

As an aside, it was noted which were the most common aspects that participants stated that they referred to the most when making the judgements and decisions they were questioned about. Table 4.9 below shows that there was considerable variation in responses but that generally, sleep duration was the most important aspect considered when making judgements and decisions about sleep. Experiencing trouble staying awake during the day was the least important.

Table 4.9.
The Number of Participants Stating Each Sleep Aspect as the Aspect they referred to the Most When Making Judgements About their Sleep.

Judgement	Duration	Latency	Disturbance	Daytime Sleepiness	Non-restorative Sleep
Worry	211 (32.7%)	99 (15.3%)	130 (20.2%)	70 (10.9%)	135 (20.9%)
Distress	171 (26.5%)	100 (15.5%)	136 (21.1%)	84 (13.0%)	154 (23.9%)
Sleep Problem	190 (29.5%)	114 (17.7%)	147 (22.8%)	50 (7.8%)	144 (22.3%)
Sleep Quality	237 (36.7%)	62 (9.6%)	141 (21.9%)	41 (6.4%)	164 (25.4%)
Help-seeking	101 (15.7%)	99 (15.3%)	106 (16.4%)	63 (9.8%)	104 (16.1%)

Note. $N = 645$.

Similar to previous results, Table 4.10 below shows that the most important predictor of participants' worry about sleep and distress caused by sleeping problems was their absolute experience of the aspect they referred to the most when making these judgements. The step 1 models accounted for 34% of the variance in worry ratings and 31% of the variance in distress ratings. When rank is entered in step 2a, the amount of explained variance increases significantly, but by only 1% for both models and the absolute experience remains the most important predictor of the worry and distress ratings. When distance from the average is added in step 2b, the amount of explained variance only increases significantly in the distress model (by 1%) and again, the absolute experience remains the most important predictor. Model comparison showed that there was no significant difference in the amount of variance accounted for by the step 2a and 2b models in worry, $t(642) = 1.29$, $p = .098$, or distress ratings $t(642) = 0.34$, $p = .366$. These results suggest that social comparison may not be fuelling sleep-related worry and distress as much as originally hypothesised.

Table 4.10.

Multiple Linear Regression Analyses Predicting Overall Worry about Sleep and Distress Caused by Sleeping Problems from Age, Gender, Absolute Experience of the Sleep Aspect they Based their Judgement On, Rank of this Sleep Aspect, and Distance from the Believed Average Experience of the Sleep Aspect in People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Worry					Distress				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.34***				**<.001**	.31***				**<.001**
Constant		45.67 (39.65-51.48)	3.14		.001		33.22 (27.40-39.34)	2.98		.001
Age		**-0.37 (-0.52--0.20)**	**0.08**	**-.15**	**.001**		**-0.27 (-0.43--0.12)**	**0.08**	**-.12**	**.001**
Gender		-1.75 (-5.23-1.90)	1.87	-.03	.347		0.82 (-2.60-4.19)	1.72	.02	.643
Absolute Experience		**14.79 (13.26-16.47)**	**0.87**	**.58**	**.001**		**12.88 (10.91-14.83)**	**0.94**	**.55**	**.001**
Step 2a	.01***				**<.001**	.01*				**<.001**
Constant		44.72 (38.65-50.81)	3.15		.001		32.72 (26.91-39.10)	2.98		.001
Age		**-0.35 (-0.51--0.19)**	**0.08**	**-.15**	**.001**		**-0.26 (-0.43--0.11)**	**0.08**	**-.12**	**.001**
Gender		-1.03 (-4.65-2.83)	1.89	-.02	.585		1.15 (-2.30-4.62)	1.74	.02	.517
Absolute Experience		**13.49 (11.83-15.31)**	**0.93**	**.53**	**.001**		**12.18 (10.26-14.19)**	**0.99**	**.52**	**.001**
Rank		**3.60 (1.38-5.84)**	**1.08**	**.13**	**.001**		**2.08 (0.25-4.11)**	**0.96**	**.08**	**.035**
Step 2b	.00				**<.001**	.01*				**<.001**
Constant		45.22 (39.54-51.37)	3.07		.001		32.96 (27.06-38.95)	2.94		.001
Age		**-0.36 (-0.51--0.21)**	**0.08**	**-.15**	**.001**		**-0.27 (-0.42--0.12)**	**0.08**	**-.12**	**.002**
Gender		-1.47 (-5.09-2.35)	1.91	-.03	.453		1.18 (-2.39-4.86)	1.82	.02	.512
Absolute Experience		**12.17 (8.92-15.28)**	**1.63**	**.47**	**.001**		**9.62 (6.55-12.92)**	**1.54**	**.41**	**.001**
Distance from the Average		3.02 (-0.30-6.49)	1.62	.12	.067		**3.78 (1.12-6.40)**	**1.46**	**.16**	**.014**

*Note. N = 645, CI = confidence interval. Significant models and predictors are highlighted in bold. *p<.05, ***p<.001*

Aim 3

In this section, participants judgements about whether they think they have a sleep problem (measured on a 0 to 100 scale from "definitely not" to "definitely yes", *M* = 45.17, *SD* = 32.74), and ratings of their overall sleep quality (measured on a 0 to 100 scale from "very bad" to "very good", *M* = 53.28, *SD* = 26.89) are predicted from the absolute, rank, and distance from the average values of the aspect each participant stated that they had referred to the most when making these two judgements.

Table 4.11 below shows a very similar pattern of results to those seen thus far. Participants' absolute experience of the sleep aspect they referred to the most when judging whether they have a sleep problem and the overall quality of their sleep was the strongest predictor of these judgements in the step 1 models. These models account for 39% of the variance in both judgements. When rank is entered in step 2a, the amount of explained variance increases significantly by 2% (sleep problem model) and 4% (quality model) but absolute experience remains the most important predictor of the judgements. Similarly, when distance from the average is added in step 2b, the amount of explained variance increases significantly by 1% in both models but the absolute experience remains the most important predictor. Model comparison showed that there was no significant difference in the amount of variance accounted for by the step 2a and 2b models in judgements of whether participants thought they had a sleep problem, $t(642) = 1.50$, $p = .067$, but model 2a accounted for significantly more variance in sleep quality judgements than model 2b, $t(642) = 2.49$, $p = .007$.

The results therefore suggest that participants did take into account how their sleep compared to that of others when judging whether they had a sleep problem and the overall quality of their sleep. The findings are inconclusive as to how they did so when judging whether they had a sleep problem but suggest that they compared using rank-based strategies when judging the quality of their sleep.

Table 4.11.

Multiple Linear Regression Analyses Predicting Whether Participants think they have a Sleep Problem and Overall Quality of Sleep Ratings from Age, Gender, Absolute Experience of the Sleep Aspect they Based their Judgement On, Rank of this Sleep Aspect, and Distance from the Believed Average Experience of the Sleep Aspect in People Around the Same Age in the General Population (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Sleep Problem					Sleep Quality				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.39***				**<.001**	.39***				**<.001**
Constant		47.33 (40.38-54.11)	3.31		.001		50.25 (45.01-55.65)	2.67		.001
Age		**-0.27 (-0.41--0.12)**	**0.08**	**-.10**	**.003**		**0.19 (0.05-0.32)**	**0.07**	**.09**	**.005**
Gender		-0.10 (-4.21-3.82)	1.98	.00	.950		0.54 (-2.71-3.62)	1.65	.01	.760
Absolute Experience		**17.93 (16.16-19.81)**	**0.93**	**.62**	**.001**		**-14.91 (-16.46--13.51)**	**0.81**	**-.63**	**.001**
Step 2a	.02***				**<.001**	.04***				**<.001**
Constant		46.31 (39.59-52.79)	3.27		.001		50.80 (45.56-56.32)	2.58		.001
Age		**-0.25 (-0.39--0.10)**	**0.08**	**-.09**	**.002**		**0.17 (0.03-0.31)**	**0.07**	**.08**	**.008**
Gender		0.50 (-3.36-4.22)	1.95	.01	.798		0.11 (-2.95-3.14)	1.61	.00	.944
Absolute Experience		**16.14 (14.03-18.26)**	**1.03**	**.56**	**.001**		**-12.55 (-14.24--10.98)**	**0.91**	**-.53**	**.001**
Rank		**4.55 (2.57-6.48)**	**1.06**	**.15**	**.001**		**-5.46 (-7.37--3.58)**	**0.90**	**-.21**	**.001**
Step 2b	.01*				**<.001**	.01**				**<.001**
Constant		46.66 (40.33-53.09)	3.16		.001		50.88 (45.54-56.09)	2.57		.001
Age		**-0.25 (-0.41--0.08)**	**0.08**	**-.09**	**.002**		**0.18 (0.05-0.32)**	**0.07**	**.08**	**.014**
Gender		0.16 (-3.58-3.66)	2.06	.00	.926		0.05 (-3.19-2.93)	1.60	.00	.973
Absolute Experience		**14.18 (10.53-17.93)**	**1.92**	**.49**	**.001**		**-11.41 (-14.65--8.75)**	**1.57**	**-.48**	**.001**
Distance from the Average		**4.25 (0.72-7.75)**	**1.77**	**.15**	**.019**		**-4.03 (-7.10--0.66)**	**1.51**	**-.17**	**.008**

*Note. N = 645, CI = confidence interval. Significant models and predictors are highlighted in bold. *p<.05, **p<.01, ***p<.001*

Aim 4

The fourth aim of the study was to investigate whether people compare to others when making decisions about help-seeking for sleep problems, either directly or indirectly through sleep-related worry or distress and/or general sleep perceptions. Participants were only asked which of the five sleep aspects had influenced their help-seeking decision the most if they had actually sought help. This meant that the help-seeking absolute, rank, and distance from the average variables calculated using the aspect each participant stated that they had referred to the most when making help-seeking decisions were not complete. This aim is consequently investigated in a slightly different manner to the two previous aims. New absolute, rank, and distance from the average variables were constructed by calculating the average absolute, rank, and distance from the average standardised values across all five sleep aspects. These are referred to as the "average" absolute, rank, and distance from the average variables. Direct associations will be explored first. Tables 4.12 and 4.13 below show the results from regression analyses investigating predictors of:

1. The number of help-seeking behaviours participants engaged in to manage sleeping problems (*M* = 1.99, *SD* = 1.86),
2. Whether they had sought help at all for sleeping problems (73.6%),
3. Whether they had consulted a professional about their sleeping problems (8.2%), and
4. Whether they had taken medication for sleeping problems (30.7%).

Table 4.12.

Multiple Linear Regression Analyses Predicting the Number of Sleep Help-seeking Behaviours Participants Engaged with and Binary Regression Analyses Predicting Whether Participants Sought Help for Sleep Problems from Age, Gender and the Average Absolute, Rank, and Distance from the Average Variables (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Number of Help-seeking Behaviours					Whether Participants Sought Help					
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.21***				**<.001**	.17***				**<.001**	
Constant		2.37 (1.94-2.79)	0.19		.001			1.41 (0.77-2.08)	0.32	.001	4.10
Age		**-0.01 (-0.02-0.00)**	**0.01**	**-.08**	**.018**		-0.04	0.00 (-0.02-0.01)	0.01	.704	1.00 (0.98-1.01)
Gender		0.05 (-0.21-0.29)	0.13	.01	.733		-0.05	-0.10 (-0.51-0.31)	0.20	.612	0.91 (0.62-1.33)
Absolute Experience		**1.23 (1.01-1.43)**	**0.10**	**.45**	**.001**		**0.99**	**1.44 (1.05-1.93)**	**0.20**	**.001**	**4.22 (2.93-6.09)**
Step 2a	.01*				**<.001**	.00				**<.001**	
Constant		2.35 (1.93-2.75)	0.19		.001			1.42 (0.78-2.11)	0.32	.001	4.16
Age		**-0.01 (-0.02-0.00)**	**0.01**	**-.07**	**.018**		-0.04	0.00 (-0.02-0.01)	0.01	.702	1.00 (0.98-1.01)
Gender		0.08 (-0.19-0.33)	0.13	.02	.566		-0.06	-0.12 (-0.53-0.29)	0.20	.538	0.89 (0.61-1.31)
Absolute Experience		**1.11 (0.88-1.34)**	**0.11**	**.41**	**.001**		**0.93**	**1.36 (0.91-1.87)**	**0.23**	**.001**	**3.89 (2.63-5.75)**
Rank		**0.30 (0.06-0.52)**	**0.11**	**.10**	**.012**		0.12	0.19 (-0.14-0.51)	0.17	.247	1.21 (0.87-1.69)
Step 2b	.00				**<.001**	.01				**<.001**	
Constant		2.36 (2.00-2.77)	0.19		.001			1.44 (0.76-2.16)	0.35	.001	4.23
Age		**-0.01 (-0.02-0.00)**	**0.01**	**-.08**	**.020**		-0.04	0.00 (-0.02-0.01)	0.01	.655	1.00 (0.98-1.01)
Gender		0.06 (-0.21-0.33)	0.14	.02	.625		-0.06	-0.13 (-0.54-0.26)	0.20	.513	0.88 (0.60-1.30)
Absolute Experience		**1.07 (0.72-1.47)**	**0.18**	**.40**	**.001**		**0.79**	**1.15 (0.59-1.78)**	**0.29**	**.001**	**3.15 (1.89-5.26)**
Distance from the Average		0.19 (-0.15-0.50)	0.18	.07	.307		0.26	0.38 (-0.07-0.87)	0.24	.105	1.47 (0.91-2.36)

Note. N = 645, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. *Significant models and predictors are highlighted in bold. *p<.05, ***p<.001*

Table 4.13.

Binary Regression Analyses Predicting Whether Participants Sought Help from a Professional (Doctor, Nurse, Pharmacist, Counsellor, Psychologist or Any Other Professional) or Took Medication (Prescribed or Over-the-Counter) for Sleep Problems from Age, Gender and the Average Absolute, Rank, and Distance from the Average Variables (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Sought Help from a Professional						Took Medication					
Predictors	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)	ΔR_N^2	B*	B (95% CI)	Std. Error	*p*	Odds Ratio (95% CI)
Step 1	.20***				**<.001**		.17***				**<.001**	
Constant			-3.70 (-4.92--2.57)	0.60	.001	0.03			-1.36 (-1.94--0.78)	0.31	.001	0.26
Age		**0.29**	**0.02 (0.00-0.05)**	**0.01**	**.041**	**1.03 (1.00-1.05)**		**0.21**	**0.02 (0.00-0.03)**	**0.01**	**.028**	**1.02 (1.00-1.03)**
Gender		-0.03	-0.06 (-0.72-0.57)	0.33	.849	0.94 (0.51-1.75)		-0.14	-0.29 (-0.68-0.11)	0.19	.119	0.75 (0.52-1.08)
Symptom Occurrence		**0.96**	**1.41 (1.02-1.90)**	**0.23**	**.001**	**4.09 (2.69-6.23)**		**0.75**	**1.10 (0.84-1.40)**	**0.14**	**.001**	**3.00 (2.28-3.95)**
Step 2a	.05***				**<.001**		.01				**<.001**	
Constant			-3.95 (-5.15--2.78)	0.63	.001	0.02			-1.37 (-1.94--0.80)	0.31	.001	0.26
Age		**0.34**	**0.03 (0.00-0.05)**	**0.01**	**.021**	**1.03 (1.01-1.05)**		**0.22**	**0.02 (0.00-0.04)**	**0.01**	**.025**	**1.02 (1.00-1.03)**
Gender		-0.09	-0.18 (-0.84-0.42)	0.35	.594	0.83 (0.44-1.58)		-0.15	-0.31 (-0.72-0.09)	0.19	.099	0.73 (0.51-1.06)
Symptom Occurrence		**0.77**	**1.13 (0.65-1.72)**	**0.25**	**.001**	**3.09 (1.97-4.85)**		**0.69**	**1.01 (0.73-1.33)**	**0.15**	**.001**	**2.74 (2.04-3.68)**
Rank		**0.67**	**1.08 (0.39-1.85)**	**0.34**	**.003**	**2.93 (1.67-5.15)**		0.16	0.26 (-0.10-0.63)	0.17	.117	1.30 (0.94-1.80)
Step 2b	.01				**<.001**		.00				**<.001**	
Constant			-3.74 (-4.85--2.75)	0.58	.001	0.02			-1.35 (-1.91--0.78)	0.30	.001	0.26
Age		**0.32**	**0.03 (0.00-0.05)**	**0.01**	**.028**	**1.03 (1.00-1.05)**		**0.21**	**0.02 (0.00-0.03)**	**0.01**	**.021**	**1.02 (1.00-1.03)**
Gender		-0.06	-0.13 (-0.78-0.50)	0.33	.681	0.88 (0.47-1.65)		-0.15	-0.31 (-0.69-0.06)	0.19	.106	0.74 (0.51-1.06)
Symptom Occurrence		**0.59**	**0.86 (0.03-1.56)**	**0.39**	**.027**	**2.36 (1.08-5.19)**		**0.63**	**0.92 (0.45-1.36)**	**0.24**	**.001**	**2.52 (1.58-3.99)**
Distance from the Average		0.44	0.63 (-0.26-1.62)	0.45	.157	1.88 (0.87-4.09)		0.15	0.21 (-0.26-0.76)	0.25	.381	1.24 (0.79-1.95)

Note. N = 645, CI = confidence interval, B* = semi-standardised coefficient, R_N^2 = Nagelkerke Pseudo-R^2. Significant models and predictors are highlighted in bold. ***p < .001.

Total number of help-seeking behaviours: The step 1 model in Table 4.12 accounts for 21% of the variance in the total number of help-seeking behaviours participants engaged with in order to manage sleep problems. Consistent with previous results, the absolute variable was the strongest predictor in the step 1 model. When the rank variable was entered in step 2a, the amount of variance accounted for by the model increases slightly but significantly by 1%. Although participants' average believed rank is a significant predictor of the total number of help-seeking behaviours engaged with, its effect is significantly smaller than that of participants' average absolute experience of the five sleep aspects. No effect is seen when the distance from the average variable is added in step 2b. There was no significant difference in the amount of variance accounted for by the step 2a and 2b models, $t(642) = 1.03$, $p = .152$, confirming that the effect of rank is very small.

Whether sought help at all: Again, Table 4.12 shows that the absolute variable is the most important predictors of whether participants sought help at all in all steps of the model. The step 1 model accounts for 17% of the variance in this decision. This time when the rank variable is entered in step 2a, no effect is seen. Again, no effect is seen when the distance from the average variable is added in step 2b.

Whether sought help from a professional: A slightly stronger rank effect is seen in Table 4.13. The step 1 model accounts for 20% of the variance in decisions to seek help from a professional with the absolute variable being the most important predictor of this outcome. This amount of explained variance increases significantly by 5% when the rank variable is added in step 2a but the absolute variable remains the most important predictor in the model. Again, no effect is seen when the distance from the average variable is added in step 2b. Although the amount of explained variance increases by 1%, this is not a significant increase. This time there was strong evidence that the step 2a model fitted the data better than the step 2b model (differences in -2LL: -12.09). However, these results should be interpreted with caution as consulting a professional was rare (only 8% of

participants did so) meaning that the maximum likelihood estimation of the model is likely to suffer from small-sample bias (King & Zeng, 2001). Inspection of the classification tables showed that only three extra people were correctly classified when rank was added to the step 1 model; this resulted in just a 0.4% increase in overall classification accuracy. In general, the ability of the step 2a model to correctly classify participants who sought help from a professional was extremely low (9.4% correctly classified).

Whether took medication: As can be seen in Table 4.13, no comparison effects were seen at all for this outcome. The absolute variable was the most important predictor of whether participants took medication for sleep problems. The step one models accounted for 17% of the variance in these decisions.

In summary, the results provide little evidence of any association between beliefs about how one's sleep compares to others and help-seeking. It therefore did not make sense to investigate any indirect effects as planned when there was so little support for the existence of direct effects.

Aim 5

The fifth aim of the study was to investigate how people compare to others when making sleep-related judgements and decisions about help-seeking, namely whether they use rank or average-based comparisons as consistent with DbS and ALT respectively. Overall, the results reported above suggest that when people compare to others to make these judgements and decisions they tend to do so using rank-based strategies. Out of the 18 models reported above, rank was a significant predictor in 16 models whereas distance from the average was a significant predictor in only 7 (these 7 models were ones where rank was also a significant predictor; in the 4 models where rank was not a significant predictor distance from the average was not either). However, the rank effects were generally small - on average the rank variables, when significant, accounted for just 2.4% of explained variance in the outcomes measured (range: 1-7%). The rank effects were only

significantly greater than the effects of the distance from the average variables in 3 out of the 7 models where they were both significant predictors and 9 out of the 16 models where comparison effects were seen. However, the distance from the average effects were never stronger than the rank effects, there was simply no significant difference between the two. The distance from the average variables never accounted for any more than 1% of the variance in any of the outcomes measured.

Aim 6

As there was so little evidence of a relationship between participants' beliefs about how their sleep compares to that of others and help-seeking, associations between these beliefs and help-seeking accuracy were not investigated.

Discussion

Overview

This study primarily investigated whether there was any association between people's beliefs about how much their sleep differed from that of others and sleep-related worry and distress. Previous research has shown that dysfunctional beliefs play a role in the maintenance and exacerbation of sleep disorders, as they are part of the negative cognitive activity that causes heightened arousal and emotional distress that leads to sleep disturbance (Harvey, 2002). It was hypothesised that beliefs about how one's sleep compares to that of others may be one type of dysfunctional belief involved in this process. This had not been investigated previously, possibly because a standardised measure of dysfunctional beliefs (the DBAS) has been used extensively in both research investigating dysfunctional beliefs and in therapy to identify and address such beliefs. This scale does not include items that specifically measure beliefs about how much one's sleep differs from that of others, i.e., perceived sleep norms. However, the results reported here provided little evidence that such beliefs were associated with either worry related to specific aspects of sleep or overall worry about sleep and distress caused by sleep problems.

Out of these outcomes, the largest comparison effects were seen when participants were asked about how worried they were about five specific aspects of their sleep. These were the average number of hours sleep they got (sleep duration) and time they took to get to sleep over the last 90 days (sleep latency) and the number of days in the last 90 that they had trouble sleeping because of waking up throughout the night (sleep disturbance), trouble staying awake during the day (daytime sleepiness) and felt that their sleep was not refreshing (non-restorative sleep). The rank comparison variable for each aspect (i.e., the percentage of people about the participant's age in the general population that each participant believed slept for more hours, took less time to get to sleep and who had fewer days where they had trouble sleeping because of night awakenings, staying awake during the day and not feeling refreshed after sleep) accounted for 2%, 5%, 2%, 2% and 1% of the variance in judgements of how worried participants were about each of these aspects respectively. The distance from the average comparison variables (i.e., the difference between the participant's absolute experience of each aspect and what they believed to be the average experience of each aspect in people about their age from the general population) accounted for 1% of the variance in worry about sleep duration and latency judgements and no significant amount of the variance in worry about sleep disturbance, daytime sleepiness and non-restorative sleep judgements. For sleep latency, disturbance, and daytime sleepiness the amount of variance in worry judgements accounted for by the rank variables was significantly greater than that accounted for by the distance from the average variables. These effects are small and suggest only a weak association between worry related to specific aspects of sleep and beliefs about how experiences of these sleep aspects compare to others - specifically, where participants believed their experience of the sleep aspect ranked in comparison to others.

The effects seen when participants' rank and distance from the average values for the sleep aspect they referred to the most when making the overall worry and distress

judgements were used to predict these judgements were even smaller. The rank variables accounted for 1% of the variance in worry and distress judgements and the distance from the average variables accounted for no significant amount of the variance in worry judgements and 1% of the variance in distress judgements. When similar comparison variables were used to predict participants' ratings of the overall quality of their sleep and their beliefs about whether they think have a sleep problem, the effects seen were again significant but small. The rank variables accounted for 2% and 4% of the variance in sleep problem and quality judgements respectively whereas the distance from the average variables accounted for only 1% of the variance in these judgements (the 3% difference in the amount of explained variance in quality ratings between the two comparison variables was significant).

This study also investigated whether there was any association between people's beliefs about how their sleep compares to that of others and decisions regarding seeking help for sleep problems. Overall, there was very little evidence to suggest that such relationships existed. The distance from the average variable used in these analyses was not a significant predictor of any of the help-seeking outcomes (total number of help-seeking behaviours, whether participants sought any kind of help, whether they consulted a professional and whether they took medication). The rank variable only significantly predicted the total number of help-seeking behaviours participants engaged in and whether they consulted a professional for sleep problems. This variable accounted for 1% and 5% of the variance in these decisions, however, the latter estimate is likely to be inaccurate due to small-sample bias.

Overall, these results were not as predicted and suggest that, of the independent variables measured, participants' absolute sleep experience was the most important predictor of sleep-related worry and distress, quality ratings, beliefs about the presence of a sleep disorder and help-seeking. Participants' beliefs about how this experience

compares to that of others played a very small role in their judgement and decision-making. These findings suggest that inaccurate perceptions of how much one's sleep differs from that of others may not be fruitful targets for CBT-I or education campaigns aiming to make people feel better and worry less about their sleep.

Study Limitations

There were some limitations to the study that may have contributed to the findings. The main outcomes relating to judgements about sleep (the general worry, distress, quality, and sleep problem outcomes) were predicted using variables constructed from the absolute, rank, and distance from the average values relating to the sleep aspect that each participant stated that they referred to the most when making the judgements. The outcomes relating to decisions made about sleep-related help-seeking were predicted by participants' average absolute, rank, and distance from the average values across all five aspects of sleep. It is possible that these variables do not accurately reflect what participants were actually thinking about when making judgements and decisions about their sleep because 1) participants may have taken into account more than one aspect but not all of the aspects when making the judgement/decision and 2) these aspects did not encompass all aspects of sleep. For example, in the context of sleep disturbance participants were asked only about the amount of times they had trouble sleeping because of waking up during the night. Sleep could have been disturbed by many other factors such as noise, difficulty breathing, pain, or being too hot or too cold although it could be argued that all of these factors may ultimately cause waking during the night. Instead of using these constructed variables, it may have been better to investigate general sleep-related judgements and decisions using general absolute, rank, and distance from the average measures. Such measures could have asked something like "on how many of the last 90 nights have you had problems sleeping?", "out of 100 people, how many do you think have had problems sleeping on more nights than you over the last 90 nights?" and "on average,

on how many of the last 90 nights do you think people had problems sleeping?". These measures could have been included alongside the specific absolute, rank, and average measures.

However, these limitations apply to all of the main variables of interest – the absolute, rank, and distance from the average variables. If they had a large effect on the results then it would be expected that the effects of all three of these variables would be small and this was not what was observed. Consistently, across all models and outcomes, the effects of the absolute variables were large and significantly larger than any comparison effects. This suggests that the small comparison effects seen were not due to these methodological limitations and may simply reflect the actual size of these effects.

Conclusions

To conclude, the results from this study provide very little evidence of an association between how people believe their sleep compares to that of others and sleep-related worry and distress, beliefs about having a sleep disorder, perceptions of sleep quality and help-seeking for sleep problems. This suggests that beliefs about how one's sleep differs from that of others may not be fuelling the kind of affective responses that cause, exacerbate, and maintain sleep disorder as originally hypothesised.

Chapter 5: General Health

Overview

General health is often measured in large-scale surveys by a single question asking respondents to rate their overall health on a five-point scale. Answers to these self-rated health questions are used extensively in research on population health, for policy making and resource allocation. However, research suggests that self-rated health (SRH) may not approximate actual health very well, with many studies showing that individuals frequently under- or overestimate their health when giving subjective health ratings. Although it has been suggested that this is due to people using different comparison samples when answering the question, exactly how the use of different comparison samples can lead to under- or overestimations has not previously been explained. Therefore, the main aims of this study are to investigate the extent to which people compare to others when making SRH judgements, whom they compare to, and how exactly they compare. Other aims of the study include investigating differences in participants' beliefs about others and the impact they may have on both SRH judgements and measures, and investigating the extent to which people compare their health to their own previous health as well as that of others and investigating what aspects of their health people compare.

Introduction

Self-rated health (also referred to as self-perceived health or self-assessed health) is a measure of general health that is obtained by asking individuals to rate their overall health on a four- or five-point scale typically from "poor" to "excellent" or "very good". The measure is deliberately vague; it allows people to use their own definition of 'health' when making the assessment (Snead, 2007) and therefore may capture elements of health that other, more detailed or specific questions would not (Au & Johnston, 2013).

Due to their brevity and inclusivity, SRH questions are often used in large-scale national and international surveys (such as the British Household Panel Survey in the UK, the National Health Interview Survey in the USA and the World Values Survey) to measure population health. Data from SRH questions obtained in such surveys are subsequently used (and recommended for use by the World Health Organisation [WHO, 1996] and the Euro-REVES 2 project [Robine, Jagger, & The Euro-REVES 2 Group, 2003]) to inform health-related policy such as the planning and allocation of health care resources. SRH is the most commonly used health measure in economics research (Au & Johnston, 2013), having been used to investigate relationships between health and factors such as socioeconomic status (Contoyannis, Jones, & Rice, 2004; Smith, 1999), education (Johnson, 2010; Silles, 2009) and income (Ettner, 1996; Meer, Miller, & Rosen, 2003). SRH is also included in health measures such as the SF-36 health survey instrument (Ware & Gandek, 1998) which is one of the most widely used health measures in the world (Ware et al., 2008), and is frequently used as an outcome measure in clinical trials (Fayers & Sprangers, 2002).

SRH has repeatedly been shown to predict morbidity and mortality (e.g., Benyamini & Idler, 1999; Ferraro, Farmer, & Wybraniec, 1997; Idler & Benyamini, 1997; Ford, Spallek, & Dobson, 2008; Jylhä, 2009), often more strongly than disease specific measures (e.g., Ferraro & Farmer 1999; Idler & Benyamini 1997; Idler & Kasl, 1991, Kaplan & Camacho, 1983). SRH is also a strong predictor of subsequent disability (e.g., Idler & Kasl, 1995; Kaplan, Strawbridge, Camacho, & Cohen, 1993; Mansson & Rastam, 2001) and use of medical care (e.g., van Doorslaer, Jones, & Koolman, 2004; Pinquart, 2001), functional limitations (Idler & Benyamini, 1997; Idler, Russell, & Davis, 2000) and health-related behaviour (Cott, Gignac, & Badley, 1999; Manderbacka, 1998).

However, despite its widespread use and predictive power, the validity of SRH as a measure of objective health continues to be questioned in the literature (e.g., Powdthavee, 2009). Correlations between SRH and objective health measures such as physician

assessments or the number of chronic illnesses, disorders or symptoms a person has tend to be modest (rarely greater than .30: Suls, Marco, & Tobin, 1991; between .21 and .46: Singh-Manoux et al., 2006). The discordance in self-rated and physician-rated health has been reported to range from 32% to 78% of cases (e.g., DeSalvo & Muntner, 2011: 46%; Geest, Engberg, & Lauritzen, 2011: 32%; Kjvinen, Halonen, Eronen, & Nissinen, 1998: 64%; Mellner & Lundberg, 2003: 78%; Smith & Goldman, 2011: 63%; Undén & Elofsson, 2001: 40%). Furthermore, using regression analyses with data from the German Socio-Economic Panel, Jürges (2008) shows that, when predicting mortality, the independent effects of covariates such as age, sex, marital status, and income are barely attenuated when SRH is added as a predictor variable to the model. Jürges (2008) argues that although SRH is a strong predictor of mortality in isolation, if it was a valid or unbiased measure of objective health then it should absorb many of the health risks related to these covariates therefore reducing their predictive power.

In order to determine whether SRH is a good measure of objective health, discrepancies between subjective and objective health ratings (sometimes referred to as 'reporting behaviour') have been examined in samples where the two can be compared. This research has repeatedly shown that individuals often under- or overestimate their health when giving subjective health ratings (e.g., Baron-Epel et al., 2005; Jylhä, Guralnik, Balfour, & Fried, 2001; Layes, Asada, & Kepart, 2012). It is commonly suggested that this is because individuals compare to others when making the SRH judgement and use different comparison or 'reference' groups when judging their health (e.g., Groot, 2000; Powdthavee, 2009; van der Zee, Buunk, & Sanderman, 1995). This has been referred to in the literature as 'scale of reference bias' (Groot, 2000), 'adaptation bias' (see Groot, 2000), 'cut-point shift' and 'index shift' (Lindeboom & van Doorslaer, 2004), 'state-dependent reporting bias' (Kerkhofs & Lindeboom, 1995), 'scale recalibration' (Ubel, Jankovic, Smith, Langa, & Fagerlin, 2005) and 'differential item functioning' (Hays, Morales, & Reise, 2000).

Whilst there has been much research on the role of social comparison when making SRH judgments, a number of important questions remain unanswered. How do people construct comparison groups and how exactly do they compare their health to that of others? A greater understanding of the cognitive processes underling SRH judgements may provide a more detailed explanation of why people over- and underestimate their health and how general health may be better measured. Therefore, the main aims of this study are to investigate the extent to which people compare to others when making SRH judgements, whom they compare to and how exactly they compare. Other aims of the study include investigating differences in participants' beliefs about others and the impact they may have on both SRH judgements and measures, and investigating the extent to which people compare their health to their previous health as well as that of others and investigating what aspects of their health people compare.

The rest of this introduction is divided into two sections. The first concerns the use of social comparison when making SRH judgements (the main focus of the study) and the literature on this topic is briefly reviewed. An explanation is then given as to how the current research expands on the previous before the aims of the study relating to this issue are outlined. The second section briefly reviews previous research on the aspects of health that people use when making SRH judgements, explains how this is extended in the current study, and outlines the aims of the study in regard to this topic.

Judging Self-Rated Health Using Social Comparison

Research that has explored the role of social comparison when making SRH judgements has investigated whether and how much people compare to others, whom they compare to, the extent to which social comparisons influence SRH judgements and whether people make different SRH judgements when asked to compare to different comparison groups. For example, Kaplan and Baron-Epel (2003) asked participants if they compared themselves to others after making a SRH judgement and if so to whom. Nearly

80% of participants (*n* = 258) said that they used a comparison group to evaluate their subjective health and 52% of these individuals reported that people their age were the most influential comparison group when making the SRH judgement. Other groups used for comparison were friends/acquaintances (24%), sick people in general (8%), a specific acquaintance (5%), people with similar illnesses (3%), work colleagues (2%), people that live in the same district (2%), people in the same socioeconomic class (2%) and a particular sick person known to the participant (2%). Carrieri (2012) used data from over 128,000 individuals in the Italian Health Conditions Survey to investigate the extent to which SRH judgements could be predicted by how an individual's objective health (measured through the presence or absence of one or more chronic or disabling illnesses) compared to that of others in constructed comparison groups. He found substantial evidence that individuals' SRH judgements were influenced by the health of similar others (those of the same age, socioeconomic status and who lived in the same area) and the health of people that they lived with and suggested that SRH judgements were made using the health of others as a benchmark. These effects remained when income and other socioeconomic and demographic variables were controlled for.

Both Fayers, Langston, and Robertson (2007) and Ubel et al. (2005) showed how individuals with the same objective health give different health-related quality of life (HRQL) and SRH ratings respectively depending on the comparison group used when making the rating. Fayers et al. (2007) found that the comparison group patients chose when making a HRQL judgement was not related to illness severity but that there were significant differences in mean HRQL scores across comparison groups. Those who said they compared to themselves before they were ill had significantly higher (worse) HRQL scores than those who said they compared to healthy peers. Furthermore, the same finding was seen at three different time points throughout the study, despite some participants changing their comparison group over time. Ubel et al. (2005) gave participants (all older

adults) one of three versions of the SRH question: a standard version, one asking them to compare to people their own age and one asking them to compare to 20-year-olds. They found that although there were no significant differences in the number of reported illnesses or the number of reported limitations to daily living activities across groups, participants comparing their health to that of a 20-year-old reported significantly worse health (mean rating 66 of 100) than those answering the standard SRH and same-aged comparison questions (both mean rating 73 of 100). This is perhaps an extreme example, as it may be unlikely that older adults would compare to those so much younger, but nonetheless demonstrates the effect that using different comparison groups can have on SRH ratings.

These studies provide evidence that SRH judgements are, in part, made through a comparison of one's own health to others. However, the precise cognitive mechanisms through which these comparisons are made remain untested. Studies that have simply asked participants how they made a SRH judgement (e.g., Kaplan & Baron-Epel, 2003; Krause & Jay, 1994) have found that some participants compared their own health to others but do not go as far as asking the exact process used in this comparison and subsequent evaluation or the extent to which the judgement was based on the comparison. Studies using regression models to investigate whether group comparisons predict SRH (e.g., Powdthavee, 2009; Carrieri, 2012) assume that comparisons are made to the average health of the comparison group and use this in the models.

As explained in the introduction chapter, the idea that we compare to the average is widely assumed but yet largely untested in the social comparison literature. Through applying ALT and DbS to SRH judgements, the actual mechanisms used in the comparison process can be tested for the first time. Applied to SRH judgements, ALT would suggest that people judge how well they are based on how their health compares to the average health of others in a comparison sample. DbS on the other hand would suggest that people judge

how well they are based on where their health ranks amongst the health of others in the comparison sample.

Therefore, the main aims of the current study are:

1. To measure the strength of the relationship between subjective and objective health,
2. To investigate the extent to which people compare to others and to their previous health when making SRH judgements and to further investigate who they compare to,
3. To understand how people make comparisons, i.e., whether they use rank or average-based strategies, and
4. To investigate how beliefs about others may differ and the impact this may have on both SRH judgements and measures.

As mentioned earlier, although the main aim of the study is to investigate the extent to which people compare to others when making SRH judgements and how they make these comparisons, the aspects of health that people use when making SRH judgements are also investigated and this is explained in detail in the section below.

Aspects of Health Used in Self-Rated Health Judgements

Jylhä (2009) proposes a model that describes the process involved in evaluating one's health and making the SRH judgement. The process involves three stages of evaluation. Firstly, the individual must decide what they consider "health" to be and what aspects of health (e.g., presence or absence of symptoms/conditions/illnesses, engagement with healthy and/or non-healthy behaviours, physical functioning etc.) they should use when making the SRH judgement. Once this is determined, the individual then considers how to take each aspect into account and how to amalgamate them in order to make an overall judgement about their health. Finally, the individual has to decide which of the given scale answers their health assessment matches best. Jylhä (2009) notes that all of the stages are influenced by contextual factors – individuals do not consider aspects of their health abstractly, they do so within different contexts such as how much their health

impacts on their daily living or how their health currently compares to their previous health or the health of others.

Studies investigating which aspects of health are used to make the SRH judgment have directly asked people which aspects of their health they brought to mind whilst making the judgement (e.g., Kaplan & Baron-Epel, 2003; Krause & Jay, 1994) or by using regression analysis to see the extent to which certain aspects predict SRH (e.g., Au & Johnston, 2013; Mavaddat et al., 2011; Schüz, Wurm, Schöllgen, & Tesch-Römer, 2011). The latter is often done using standardised measures or subscales of measures (such as the SF-36) that measure particular aspects of health such as physical health, mental health, and physical functioning. These studies have shown that individuals use a wide range of health aspects when making the SRH judgement such as the presence or absence of physical health problems or illnesses, engagement with positive or negative health behaviours (such as taking regular exercise, eating healthily, smoking and drinking alcohol), physical functioning, general physical condition, mental health and well-being, and energy/tiredness.

There is some debate as to which of these aspects is most frequently used or which is the most important aspect used when making the SRH judgment, but generally individuals refer to aspects of their physical health rather than their mental health. Functional status, vitality, and presence of illness or disease are most commonly used when making the judgement (Au & Johnston, 2013; Krause & Jay, 1994; Mavaddat et al., 2011; Powdthavee & van den Berg, 2011; Schüz, et al., 2011). Kaplan and Baron-Epel (2003) asked participants to rate how influential different aspects of health were when making the SRH judgement. The most influential factor across all participants was general feeling; pain and difficulty in performing certain activities were also important. Differences in the use of aspects have also been found across ages, education groups, and race and between those in good and bad health. Krause and Jay (1994) showed that younger people were more

likely than older people to refer to health behaviours when making the SRH judgement and older people were more likely than younger people to refer to health problems. Participants who went to university were less likely than those who did not to refer to health behaviours. They also found that Hispanic and Black people were more likely than White people to refer to health problems when judging their health and that White people were more likely than Hispanic and Black people to refer to their general physical functioning. There is also evidence that older people in poor health base their ratings on health problems or limitations whereas those in good health tend to base their ratings more on health behaviours (Benyamini, Idler, Leventhal, & Leventhal, 2000; Benyamini, Leventhal, & Leventhal, 1999; 2003). Using data from all their participants (both old and young), Kaplan and Baron-Epel (2003) found that participants in poor health rated pain, tiredness and current medication intake as being more influential than those in good health whereas those in good health rated difficulty in performing activities as being more influential.

Whilst much is known about the aspects of health that people bring to mind when making the SRH judgement, less is known about how these aspects are actually taken into account (Jylhä, 2009), i.e., the second stage of Jylhä' s (2009) model. As outlined above, it is hypothesised that people sometimes make SRH judgements using a comparative process and therefore it is hypothesised that aspects of health are brought to mind and then compared, either to other people's experiences of the aspect or the individual's previous experiences of the aspect. For example, an individual may judge how good their health is by considering the number of cigarettes they smoke and they may do this by comparing this number to the number that their friends and family smoke. Another individual may use the current state of a longstanding illness to make the judgement and may compare this to how they were 3 years ago in order to do so. Therefore, this study will also investigate

which aspects of their health individuals compare to others and their previous experiences. The aims of the second part of the study are as follows:

5. To further understand what aspects of their health people base SRH judgements on and to explore any group differences,
6. To understand what aspects of their health people use specifically to compare to others and to themselves previously, and
7. To see whether people use different strategies when making the SRH judgement depending on which aspect of their health they use when doing so.

Method

Participants

Participants (*N* = 643) were recruited through Amazon MTurk and had a mean age of 37.1 years (*SD* = 12.7, range = 18-83 years), were predominantly White (72%, Asian = 9%, Black = 8%, Hispanic/Latino = 6%, other = 5%), and 57% were female. Participants were mainly university educated (64%, high school diploma = 33%, some high school = 2%, rather not say = 1%), were all resident in the U.S. and completed the study from 49 different states. Participants received $0.75 on completion of the study which took 15-20 minutes to complete.

Design and procedure

The questionnaire that participants completed in this study was designed and hosted using Qualtrics. Participants were asked the following:

Self-rated health: SRH was measured through the following question: "In general, would you say your health is: excellent, very good, good, fair, or poor". This version was taken from the SF-36 health survey instrument (Ware & Gandek, 1998). Participants answered using a slider scale that went from 0 to 100 and which was visibly divided into 5 sections with "poor" encompassing an answer between 0 and 20, "fair" between 20 and 40 and so on.

How they made the SRH judgement: Participants were then asked "in the form of a percentage where 0% is not at all and 100% is completely, when you answered the previous question about your general health, to what extent did you: A) Compare yourself to other people in order to answer the question? B) Compare your current health with your previous health in order to answer the question? C) Base your answer on something else completely?". The question was designed such that their answers to all parts needed to add up to 100%.

Rank amongst comparison groups: Next, participants stated the percentage of people in nine different comparison groups (friends, family, work colleagues, same age ±3 years, more than 3 years younger, more than 3 years older, people who live near, general population of the USA, people with similar health conditions) that were in better health than them. This was done using slider scales for each group that were labelled from 0 to 100 with 0 labelled also as "none in better health than you" and 100 labelled also as "all in better health than you". This gave a measure of each participant's rank within the different comparison groups.

Average health of comparison groups*:* Using the same slider design as the SRH question, participants were asked what they though the average health of people from each of the nine comparison groups was. The presentation order of these questions and the rank elicitation questions above was counterbalanced so that half of the participants saw the rank questions first and half saw the average questions first.

At this stage, participants saw further questions regarding comparisons they had made when answering the SRH questions but only if they stated that they either compared to others and/or compared to themselves previously.

If participants said that they compared to *others* then they were also asked about:

1. **Their most important comparison group:** Participants were asked which of the nine comparison groups they compared to the most when answering the SRH question.

2. **The extent to which they used each comparison group**: Then participants were asked the extent to which (if at all) they compared their health to others in each of the nine comparison groups in order to answer the SRH question. This was answered using a slider scale that went from 0 to 100 which was visibly divided into 4 sections: “not at all” (0-25), “a little” (25-50), “some” (50-75), and “a lot” (75-100).
3. **The extent to which they compared aspects of their health to others:** Using the same slider scale as above, participants were finally asked about the extent to which (if at all) they compared 13 different aspects of their health to other people’s experiences of these aspects when making the SRH judgement. These aspects were: Presence of health problems, absence of health problems, physical functioning, general physical condition, energy, engagement with healthy behaviours such as regular exercise, engagement with unhealthy behaviours such as smoking, mental health, general feeling, tiredness, pain, medication you are on, treatments you receive. This list was an amalgamation of aspects of health that Krause and Jay (1994) and Kaplan and Baron-Epel (2003) had found important when evaluating SRH.

If participants said that they compared to *themselves previously* when answering the SRH question then they were also asked about:

1. **The extent to which they compared aspects of their health to their previous health:** Using the same slider scale as the two questions above, participants were asked the extent to which (if at all) they compared their current experience of the 13 different aspects of health outlined above to their previous experiences of these aspects.

All participants were then asked about:

Their most important health aspect: Participants were asked which of the 13 health aspects above was the most important when answering the SRH question.

Objective health: As it was not possible to access participants’ medical records or clinically examine participants, the self-reported number of symptoms/chronic

conditions/illnesses that participants had/were experiencing was used as a proxy measure of objective health. This follows the procedure used in other studies investigating the relationship between objective and subjective health (e.g., Groot, 2000; Powdthavee, 2009; Ubel et al., 2005; van der Zee & Buunk, 1995). Participants were asked both how many symptoms from a list of 25 they had experienced in the last six months (if any) and whether a medical professional had ever said that they had any of 17 illnesses/chronic conditions. The symptoms on the checklist were those shown by McAteer et al. (2011) to be the most commonly occurring within the general population: back pain, blood in stool, chest pain, cold or flu symptoms, constipation, cough, coughing up blood, diarrhoea, difficulty sleeping, dizziness, fainting, feeling depressed, feeling tired/run down, headaches, indigestion/heartburn, joint pain, loss of appetite, nausea/feeling sick, nervousness/anxiety, shortness of breath, sore throat, stomach/abdominal pain, unintentional weight loss, vomiting, and wheezy chest. The list of illnesses/chronic conditions was also obtained from McAteer et al. (2011) and included the following: anaemia, asthma, arthritis or rheumatism, bronchitis or emphysema, cancer, chronic liver trouble, diabetes, serious back trouble, heart trouble, high blood pressure, circulation problems, kidney or bladder problems, ulcers, allergies, multiple sclerosis, colitis, and high cholesterol.

It is possible that a self-report measure of objective health such as this will not be as accurate or reliable as “gold standard” measures such as information from medical records or clinical examinations. This is because the accuracy of self-reported illness can be affected by a number of factors including participants’ knowledge of the symptoms/conditions presented to them, their ability to recall whether they have been experienced in the period of time provided, and whether or not they are willing to report their experience (Goldman, Lin, Weinstein, & Lin, 2003). However, research has shown that for many of the symptoms/conditions listed above, agreement between self-report and

medical report (from medical records or physician diagnosis) tends to be very good (κ = 0.80 to 1.00) or good (κ = 0.60 to 0.79) suggesting that inaccurate self-reporting in the current study may not be too problematic. For example, people tend to be very good at accurately reporting whether or not they have asthma (κ = 0.83, Baumeister, Kriston, Bengel, & Härter, 2010), cancer (κ = 0.92, Baumeister et al., 2010), colitis/Chrohn's disease (κ = 0.96, Kelstrup, Juillerat, & Korzenik, 2014), and diabetes (κ = 0.90, Barber, Muller, Whitehurst, & Hay, 2010; κ = 0.87, Baumeister et al., 2010; κ = 0.85, Kriegsman, Penninx, van Eijk, Boeke, & Deeg, 1996; κ = 0.80, Muggah, Graves, Bennett, & Manuel, 2013; κ = 0.76, Okura, Urban, Mahoney, Jacobsen, & Rodeheffer, 2004) and good at accurately reporting bronchitis (κ = 0.74, Baumeister et al., 2010) and high blood pressure (κ = 0.75, Okura et al., 2004; κ = 0.67, Barber et al., 2010; κ = 0.66, Muggah et al., 2013). People are moderately accurate at reporting the presence of arthritis (κ = 0.53, Baumeister et al., 2010) and depression (κ = 0.62, Zimmerman, Coryell, Wilson, & Corenthal, 1986; κ = 0.50, Stuart, Pasco, Jacka, Brennan, Berk, & Williams, 2014). Although Barbara, Loeb, Dolovich, Brazil, and Russell (2012) found only poor to fair agreement between self-report and medical report of symptoms of respiratory illness (cough κ = 0.41, headache κ = 0.21, muscle ache κ = 0.21, sore throat κ = 0.19, fatigue κ = 0.13), total percentage agreement (positive and negative) ranged from 61% to 84% (cough = 72%, headache = 81%, muscle ache = 88%, sore throat = 61%, fatigue = 84%).

Demographic characteristics: Age, gender, level of education, ethnicity, income bracket and the state in which participants lived were also collected.

Results

In this section, the results are again reported in relation to each of the aims of the study. In cases where group differences were tested participants was split into three age groups: under 30 (*n* = 241), 30-40 (*n* = 163), and over 40 (*n* = 239). The groups were split in this manner to ensure the most even distribution of participants across groups. As 97% of

participants had either a high school diploma or a university education, differences across education groups were investigated between participants in these two groups only. As 72% of participants were White and the second largest ethnicity group comprised only 9% of participants, group differences were investigated between White and non-White participants only. Differences in reporting behaviour across SRH groups were also investigated. As previous studies had split participants into two SRH groups (good and bad), the same was done here. As there were only 33 participants who stated that their health was poor, participants were split into the following groups: Good health (those who responded either excellent, very good or good to the SRH question) and poor health (those who responded either fair or poor to the SRH question).

Before any analysis took place, participants' answers to the questions eliciting their rank within the comparison groups were inverted so that greater values indicated that they believed they were in better health than others. For example, if the participant stated that only 30% of friends were in better health than them then their rank value became 70. This was done so that the SRH, rank, and the average health of the comparison group variables were all on a scale that went from poor to excellent health. For all statistical tests, an alpha level of .05 was used.

Aim 1

The first aim was to measure the strength of the relationship between subjective and objective health. The number of symptoms experienced and the number of chronic illnesses that each individual had were added together to create an objective health measure that had greater variability and encompassed a wide variety of health aspects[8]. The resulting measure had a higher correlation with subjective health (measured from excellent to poor: $r = .39$, $p < .001$) than the number of symptoms ($r = .36$, $p < .001$) or

[8] A similar objective health measure was calculated by transforming the number of symptoms and conditions to *z*-scores first and then adding the two together. The analyses for Aim 1 and 3 were conducted again using this measure and all of the results reported were exactly the same.

number of chronic illnesses (r = .30, p < .001) alone. These correlations are similar to the average correlation of r = .30 reported by Suls et al. (1991). They are also in line with the findings of Singh-Manoux et al. (2006), who report correlations of r = .28, r = .36 and r = .42 between SRH and the number of longstanding illnesses, number of recurring health problems and the number of symptoms experienced in the last 14 days respectively. Interestingly, the correlation between SRH and objective health was significantly stronger for participants who stated that they did not compare to others when making the SRH judgement (r = .51, p < .001) than for those that did (r = .35, p < .001), z = 1.80, p = 0.036, although these groups were quite unequal in numbers (n = 102 and n = 541 respectively).

Aim 2

The second aim was to investigate the extent to which people compare to others and themselves previously when making SRH judgements and to further investigate who they compare to. When directly asking participants how they made the SRH judgement, 84% of participants (n = 541) said that they compared to others to some extent when making the judgement, 90% (n = 579) said that they compared their health now to how it was previously to some extent when making the judgement and 65% (n = 417) said they based their answer to some extent on something other than comparison to others or themselves previously. The amount of people that stated that they compared to others when making the SRH judgement (84%) is in line with previous findings by Kaplan and Baron-Epel (2003; 80%).

Across all participants, the average split of answers to the question asking how they had made the SRH judgement was as follows: by comparing their health to that of others: 37% (SD = 26%), by comparing their current health to their previous health: 43% (SD = 28%), by something other than the above: 20% (SD = 24%). On average, participants who stated that they used social comparison said they based 44% (SD = 22%) of the SRH judgement on how their health compared to that of others. Participants who stated that

they compared to their previous health stated that, on average, they based 47% (*SD* = 26%) of the SRH judgement on how their current health compared to their previous health.

There were no significant differences in the extent to which participants based their SRH judgements on how their health compared to others across gender, $t(536) = 1.48$, $p = .140$ (two-tailed), or age groups, $F(2,538) = 0.04$, $p = .961$. However, the extent to which participants based the SRH judgement on how their current health compares to their previous health differed across gender, $t(574) = 2.38$, $p = .018$ (two-tailed), and age groups, $F(2,576) = 4.28$, $p = .014$. On average, women based a higher percentage of their SRH judgement on how their health currently compared to their previous health than men (women: *M* = 50%, *SD* = 28%; men: *M* = 44%, *SD* = 23%). LSD comparisons revealed that, on average, people who were over 40 based a higher percentage of their SRH judgement on how their health currently compared to it previously than those who were under 30 (over 40: *M* = 51%, *SD* = 27%; under 30: *M* = 44%, *SD* = 25%). There were no significant differences between the under 30 and 30-40 age group or the 30-40 and over 40 age group.

The comparison group that participants stated that they used the most when comparing to others in order to make the SRH judgement was friends (23.0%), followed by family (19.3%), people within 3 years of age (18.4%), the general population (13.8%), work colleagues (5.0%), people who live nearby (1.9%), people with similar health conditions (1.9%), people who were more than 3 years younger (0.6%), and people more than three years older (0.3%).

Group differences in the comparison group used most were also examined using chi-square tests and *z*-tests (Bonferroni corrected for multiple comparisons) to examine differences in the proportions of group members selecting each comparison group as the one they compared to most when making the SRH judgement:

Gender: Gender was not significantly associated with choice of comparison group, χ^2 (8) = 10.34, *p* = .242.

Age: Age group was significantly associated with choice of comparison group, χ^2 (16) = 38.77, *p* = .001. There were significant differences between the youngest (<30) and oldest (>40) age groups in the proportion of participants selecting three of the comparison groups as their most important when making the SRH judgement. A significantly higher proportion of young people (28.1%) chose family as the group they compared to the most than of participants in the oldest age group (17.2%). The difference between the youngest and middle (30-40) and middle and older age groups was nonsignificant. In the opposite direction, a significantly greater proportion of participants in the oldest age group chose both people around their age and others with the same health conditions (28.6% and 4.7%% respectively) as their most important comparison group than of participants in the youngest age group (16.2% and 0% respectively). The difference between the youngest and middle (30-40) and middle and older age groups was nonsignificant.

Education: There was no significant association between education and choice of the most important comparison group. However, there were significant differences in the proportion of participants selecting two of the comparison groups as the ones they used most across education groups. A significantly greater proportion of participants who went to university (7.7%) chose people they work with as their most important comparison group than of participants who did not (2.9%). A significantly higher proportion of participants who did not go to university (4%) chose people who lived near them as their most important comparison groups than of those who did (1.1%).

Ethnicity: There was no significant association between ethnicity and choice of the most important comparison group. However, there was a significant difference in importance across ethnicity groups present for one of the groups; a greater proportion

(19.4%) of White participants selected the general population as their most important comparison group than of non-White participants (9.4%).

Income: Income group was significantly associated with choice of comparison group, χ^2 (32) = 50.66, p = .019. The proportion of participants selecting work colleagues as their most important comparison group was significantly greater in the highest income group (>$65,000: 12.8%) than in the lowest income group (<$14,999: 1.1%). Although the differences between the remaining income groups were nonsignificant, there was a general trend such that as income increased so did the tendency to use the health of work colleagues as a benchmark to judge one's own health.

Self-rated health: Self-rated health group was significantly associated with choice of comparison group, χ^2 (8) = 23.35, p = .003. A greater proportion of participants in good health selected family and work colleagues as their most important comparison group (24.5% and 6.8% respectively) than of participants in poor health (14.5% and 1.2% respectively). On the other hand, a greater proportion of participants in poor health chose people around their age and the general population as their most important comparison group (39.9% and 24.1% respectively) than of those in good health (19.4% and 15.1% respectively). These findings are in line with those reported by Kaplan and Baron-Epel (2003) who found that participants in good health tended to compare their health to friends whilst those in poor health compared to people their age.

Aim 3

The third aim was to investigate how people compare to others when judging their health - specifically whether they use rank- or average-based strategies. In order to do this, four different measures of social comparison (two rank-based and two average-based) were calculated and used in regression analyses predicting SRH judgements. The first set of social comparison variables, "Rank A" and "Average A", were compiled using participants' answers to the question asking 1) which comparison group they compared to the most

when answering the SRH question, 2) their believed ranked position within each of the comparison groups and 3) their believed average health of each of the comparison groups. Rank A is the participant's believed rank within the comparison group they compared to the most when making the SRH judgement and Average A is the believed average health of the comparison group they compared to the most. The second set of social comparison variables, "Rank B" and "Average B", were calculated as follows in order to include all of the available information on each participant's comparison behaviours:

$$\text{Rank B} = \sum_{g=1}^{9} Rg * Eg \quad (7)$$

$$\text{Average B} = \sum_{g=1}^{9} Ag * Eg \quad (8)$$

Where g is the comparison group (of which there are nine), Rg is the participant's rank within the comparison group, Eg is the extent to which the participant compares to the comparison group (percentage) when making the SRH judgement divided by the sum of the extent to which they compare to each comparison group, and Ag is the average health of the comparison group[9].

Age and gender are included in the regression models in order to control for any effects these covariates may have. Objective health (the number of symptoms experienced plus the number of chronic illnesses that each participant had) is also included in order to ascertain how well the rank and average variables predict SRH in comparison to objective health. Table 5.1 below shows the correlations between the continuous variables used in the regression analyses using data only from participants who stated that they compared to others to some extent.

[9] Note that as objective health and the average health of others in the comparison group were measured on different scales it was not possible to calculate how much participants' objective health differed from that of the average health of others. Therefore, the average health of others is used in analyses instead.

Table 5.1.
Pearson Correlations between Main Variables.

	SRH	Age	Objective Health	Rank A	Average A	Rank B	Average B
SRH	1	-.106*	-.353***	.304***	.175***	.325***	.198***
Age	-.106*	1	.065	.030	-.029	.005	-.003
Objective Health	-.353***	.065	1	-.220***	.018	-.268***	.041
Rank A	.304***	.030	-.220***	1	-.381***	.879***	-.332***
Average A	.175***	-.029	.018	-.381***	1	-.328***	.801***
Rank B	.325***	.005	-.268***	.879***	-.328***	1	-.388***
Average B	.198***	-.003	.041	-.332***	.801***	-.388***	1

Note. $N = 541$. *$p < .05$, two-tailed, ***$p < .001$, two tailed.

Table 5.1 above shows that the comparison variables that incorporate information about all of the participants' comparison behaviours (Rank B and Average B) correlate slightly more strongly with SRH than the comparison variables representing participants' rank within the group they compare to most (Rank A) and the average health of the group they compare to most (Average A). However, these differences are not significant, rank: $t(538) = 1.05$, $p = .148$; average: $t(538) = 0.86$, $p = .806$. The correlation between Rank A and SRH is significantly stronger than the correlation between Average A and SRH, $t(538) = 2.82$, $p = .003$, as is the correlation between Rank B and SRH compared to the one between SRH and Average B, $t(538) = 2.81$, $p = .003$. There are no significant differences in the correlations between SRH and the two rank variables and SRH and objective health, Rank A: $t(538) = 0.99$; $p = .162$; Rank B: $t(538) = 0.58$; $p = .280$. However, the correlation between SRH and objective health was significantly stronger than the correlations between SRH and both average variables, Average A: $t(538) = 3.14$; $p = .001$; Average B: $t(538) = 2.78$; $p = .003$.

Regression analyses (reported in Table 5.2 below) suggested firstly that people do compare to others when making SRH judgements: the two rank and the two average variables were significant predictors of SRH judgements. Comparison of the objective health, rank, and average beta coefficients reported in the step 2 models showed that the effects of both rank A and rank B were comparable to that of objective health – there was no significant difference between the objective health and rank beta coefficients in either

the A or B models. However, both the average variables were weaker predictors of SRH than objective health – the objective health beta coefficients are significantly larger than the average coefficients in both the A and B models. Finally, the predictive ability of the step two models was compared, using the Hotelling's t-test for dependent correlations. There were no significant differences in the variance in SRH judgements accounted for by the step 2a and 2b models when either the A or B variables were used, step two models including rank A and average A, $t(538) = 0.82$, $p = .206$, step two models including rank B and average B, $t(538) = 0.36$, $p = .361$.

Table 5.2.

Multiple Linear Regression Analyses Predicting Self-rated Health from Age, Gender, Objective Health and Both Rank and Average Variables (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Models with Rank A and Average A					*Models with Rank B and Average B*				
Predictors	ΔR^2	B (95% CI)	Std. Error	β	*p*	ΔR^2	B (95% CI)	Std. Error	β	*p*
Step 1	.13***				**< .001**	.13***				**< .001**
Constant		75.95 (70.07-81.48)	3.13		.001		75.95 (70.01-82.13)	3.06		.001
Age		-0.13 (-0.28-0.02)	0.07	-.08	.072		-0.13 (-0.28-0.00)	0.08	-.08	.071
Gender		0.62 (-2.67-4.02)	1.76	.01	.723		0.62 (-2.87-4.30)	1.79	.01	.705
Objective Health		**-1.39 (-1.68--1.09)**	**0.16**	**-.35**	**.001**		**-1.39 (-1.71--1.07)**	**0.16**	**-.35**	**.001**
Step 2a	.06***				**< .001**	.06***				**< .001**
Constant		65.14 (57.18-72.21)	3.78		.001		62.65 (55.31-70.15)	3.89		.001
Age		**-0.16 (-0.30--0.01)**	**0.07**	**-.09**	**.034**		**-0.15 (-0.29--0.02)**	**0.07**	**-.09**	**.041**
Gender		-0.03 (-3.38-3.24)	1.72	.00	.988		0.09 (-3.22-3.59)	1.75	.00	.954
Objective Health		**-1.19 (-1.52--0.86)**	**0.16**	**-.30**	**.001**		**-1.13 (-1.44--0.79)**	**0.16**	**-.28**	**.001**
Rank		**0.21 (0.12-0.28)**	**0.04**	**.24**	**.001**		**0.24 (0.16-0.33)**	**0.04**	**.25**	**.001**
Step 2b	.03***				**< .001**	.05***				**< .001**
Constant		61.05 (51.78-70.27)	4.74		.001		54.39 (43.62-65.51)	5.56		.001
Age		-0.12 (-0.25-0.01)	0.07	-.07	.093		-0.13 (-0.28-0.02)	0.07	-.07	.081
Gender		1.54 (-1.55-5.19)	1.68	.04	.356		1.70 (-1.40-4.74)	1.69	.04	.315
Objective Health		**-1.39 (-1.73--1.02)**	**0.17**	**-.35**	**.001**		**-1.41 (-1.71--1.10)**	**0.15**	**-.35**	**.001**
Average		**0.23 (0.12-0.33)**	**0.05**	**.18**	**.001**		**0.34 (0.21-0.47)**	**0.07**	**.22**	**.001**

Note. n = 541. Significant models and predictors are highlighted in bold. ****p* < .001.

Although the results provide further evidence that people compare to others when making SRH judgements, they are inconclusive as to how they do so. The correlations between the rank variables and SRH were significantly stronger than the correlations between the average variables and SRH. Furthermore, rank performed as well as objective health when predicting SRH judgements but the average was a weaker predictor. However, the model comparison showed that there was no difference in the fit of the final rank and average models to the data.

Aim 4

Aim 4 was to investigate how beliefs about others may differ and the impact that such differences may have on both SRH judgements and measures. Individual differences in beliefs about others can be explored using data from questions asking about the average health of groups. Variation in the average health of comparison groups such as friends and family would be expected as the people that make up these groups (and their health) will differ from individual to individual. There should be less variation in groups that are common to individuals such as the general population that they belong to and people in that population who are of a similar age. However, Figure 5.1 below shows considerable variation in participants' beliefs about the average health of the general population in the USA (M = 56.11, SD = 17.19, range = 0-99). Figure 5.1 also shows the variation in beliefs about the average health of other people around the same age that participants aged 32 (the modal age, n = 29) displayed (M = 67.62, SD = 16.57, range = 40-100). The impact that these differing beliefs may have on both SRH judgements and measures are explained in the discussion.

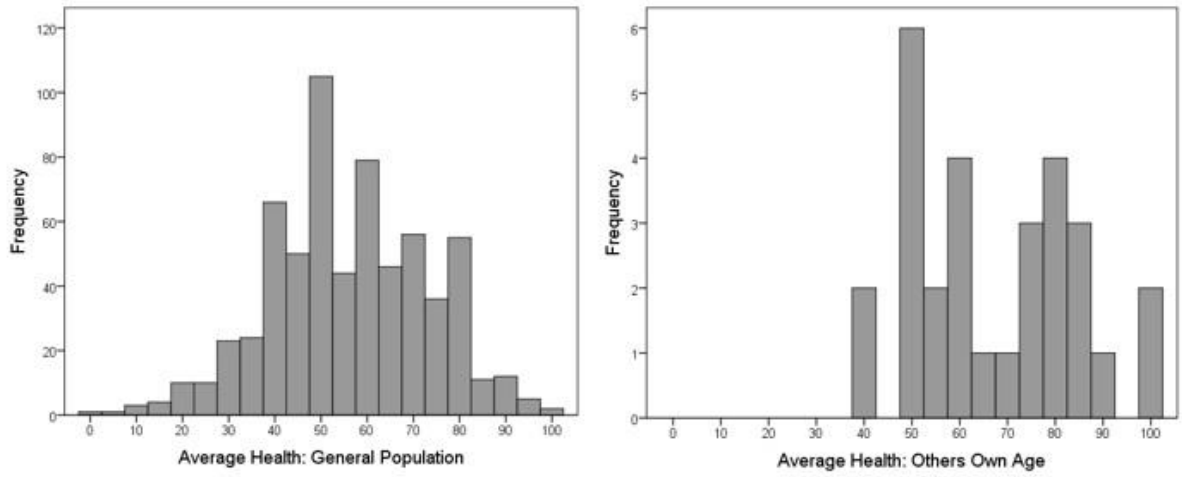

Figure 5.1. Participants' ratings of the average health of the general population in the USA (left: all participants) and of people around their age (right: participants aged 32). Scale is from poor (0) to excellent (100).

Aim 5

Aim 5 was to further understand what aspects of their health people base SRH judgements on and to explore any group differences. Participants reported that their general physical condition (28%) and the presence (23%) and absence (13%) of health problems were the most important health aspects used to evaluate their general health (see column 1 of Table 5.3 below for the full results and Figure 5.2). This is in line with previous findings (e.g., Kaplan & Baron-Epel, 2003; Krause & Jay, 1994). Also replicated is the previous finding that participants base SRH judgements on aspects of their physical health rather than their mental health - only 2% of participants stated that their mental health was the most important aspect when making the SRH judgement. Group differences in the most important health aspect chosen were also examined using chi-square tests and *z*-tests (Bonferroni corrected for multiple comparisons). These examined differences in the proportions of group members selecting each health aspect as their most important when making the SRH judgement:

Gender and income: Choice of health aspect was not significantly associated with gender, χ^2 (12) = 13.60, p = .327, or income group, χ^2 (48) = 50.88, p = .361. Krause and Jay

(1994) also found no significant association between gender and the health aspect used to make the SRH judgement.

Age: Age group was significantly associated with choice of health aspect, χ^2 (24) = 37.62, *p* = .038. There were significant differences between the youngest (<30) and oldest (>40) age groups in the proportion of participants selecting three of the aspects as their most important when making the SRH judgement. A significantly higher proportion of young people (10.4%) chose energy as their most important health aspect than of participants in the oldest age group (2.5%). The difference between the youngest and middle (30-40) and middle and older age groups was nonsignificant. In the opposite direction, a significantly greater proportion of participants in the oldest age group chose both general feeling and pain (10.5% and 6.3% respectively) as their most important health aspect than of participants in the youngest (4.6% and 1.7% respectively) age group. The difference between the youngest and middle (30-40) and middle and older age groups was nonsignificant. Krause and Jay's (1994) findings that younger people were more likely than older people to make SRH judgements based on health behaviours and that older people were more likely than younger people to make them based on health problems were not replicated.

Education: Education group was significantly associated with choice of health aspect, χ^2 (12) = 32.00, *p* = .001. A significantly greater proportion of participants who went to university (15.4%) chose the absence of health problems as their most important health aspect than of participants who did not (8%). Significantly higher proportions of participants who did not go to university chose general feeling and pain as their most important health aspect (11.3% and 6.6% respectively) than those who did (5.4% and 2.9% respectively). Aside from this there were no significant differences across education groups in the proportion of participants selecting each health aspect as their most important when making the SRH judgement.

Ethnicity: There was no significant association between ethnicity and choice of the most important health aspect. However, there was a significant difference in importance across ethnicity groups present for one of the aspects: a greater proportion (28.6%) of non-White participants selected presence of health problems as their most important aspect than of White participants (21.3%). This finding was also reported by Krause and Jay (1994).

Self-rated health: Self-rated health group was significantly associated with choice of health aspect, χ^2 (12) = 32.78, p = .001. A greater proportion of participants in good health selected general physical condition and the absence of health problems as their most important health aspect (30.1% and 13.7% respectively) than of participants in poor health (16.3% and 6.7% respectively). On the other hand, a greater proportion of participants in poor health (11.5%) chose pain as their most important health aspect when making the SRH judgement than of those in good health (2.8%). These findings are in line with those reported by Kaplan and Baron-Epel (2003) who found that participants in good health placed more importance on difficulty performing activities when making the SRH judgement whilst those in poor health focused more on pain and tiredness.

Table 5.3.
Use of Health Aspects when Making Self-rated Health Judgements and when Comparing to Others and to Previous Health.

	Most important aspect when		
	Making SRH Judgement	Comparing to Others	Comparing to Previous Health
General Physical Condition	27.8% (179)	12.5% (51)	11.5% (52)
Presence of Health Problems	23.3% (150)	5.6% (23)	8.9% (40)
Absence of Health Problems	12.6% (81)	9.3% (38)	5.5% (25)
Physical Functioning	9.0% (58)	6.1% (25)	7.8% (35)
General Feeling	7.3% (47)	8.1% (33)	11.1% (50)
Energy	6.5% (42)	11.2% (46)	9.8% (44)
Pain	4.2% (27)	10.0% (41)	8.4% (38)
Healthy Behaviours	3.6% (23)	12.0% (49)	8.0% (36)
Mental Health	2.0% (13)	6.8% (28)	7.8% (35)
Unhealthy Behaviours	1.3% (8)	5.9% (24)	8.4% (38)
Tiredness	0.9% (6)	4.2% (17)	4.2% (19)
Medication	0.9% (6)	3.4% (14)	4.0% (18)
Treatments	0.5% (3)	4.9% (20)	4.7% (21)
N	643	409	451

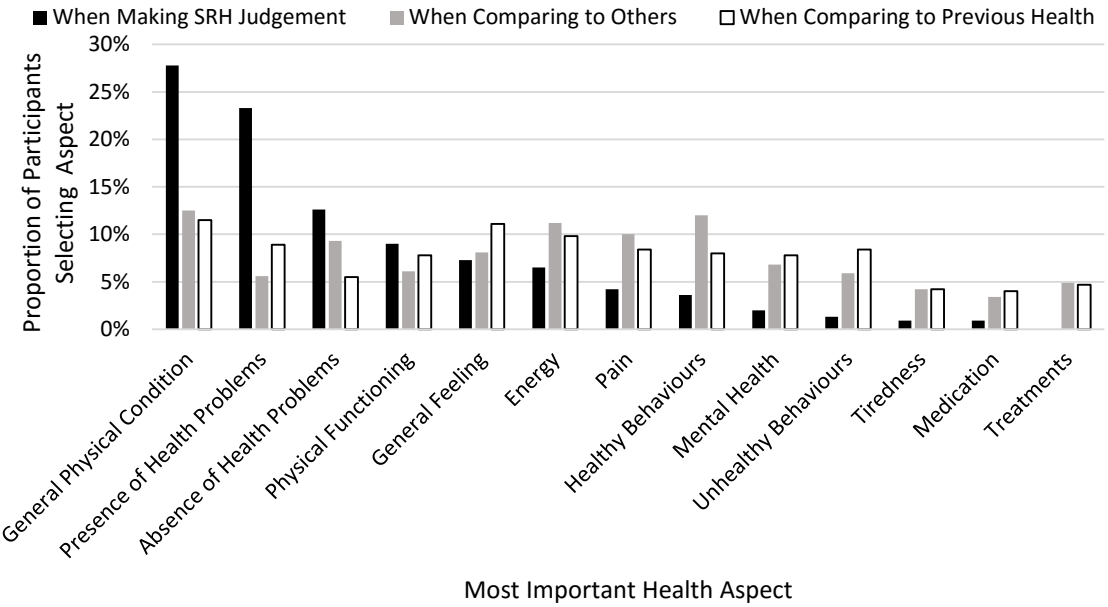

Figure 5.2. Graphical Representation of Use of Health Aspects when Making Self-rated Health Judgements and when Comparing to Others and to Previous Health.

Aim 6

Aim 6 was to understand what aspects of their health people use specifically when comparing to others or to their previous health experiences when judging their health. In order to do this, answers to the two questions asking about the extent to which participants 1) compared their current experience of the 13 different health aspects to other people's experiences of these aspects and 2) compared their current experience of the 13 aspects to their previous experiences of these aspects when making the SRH judgement were examined. Note that only participants who stated that they did compare to others or their previous health when making the SRH judgement saw these questions. The aspect that each participant gave the greatest weight to (i.e., the aspect with the highest percentage answer) was extracted as the most important aspect used when comparing to others/previous self. The proportion (and number) of participants that gave the highest weight to each aspect is shown in columns 2 and 3 of Table 5.3 and is displayed

in Figure 5.2. Participants who gave equal weights to two or more aspects (n_{others} = 132, n_{self} = 128) are not included in these figures.

When judging their general health, the health aspects that participants most commonly compared to others and to their previous health were very similar. The three aspects most commonly compared to others were general physical condition (12.5%), healthy behaviours (12%), and energy (11.2%), and to previous health were general physical condition (11.5%), general feeling (11.1%), and energy (9.8%). The three health aspects that participants least commonly compared to their previous health and others were the same. These were treatments (others: 4.9%, previous self: 4.7%), tiredness (others and previous self: 4.2%), and medication (others: 3.4%, previous self: 4%). It would appear that when the aspects are listed in rank order of the proportion of participants selecting each aspect as the most important when comparing to both others and to their previous health that there is little difference in this rank ordering between comparison type.

Aim 7

Aim 7 was to see whether people use different strategies when making the SRH judgement depending on which aspect of their health they use when doing so. Using participants' answers to the question asking the extent to which they had compared to others, compared to themselves previously or based their answer on something else completely when answering the SRH question, each participant's main answering strategy was elicited (i.e., the one with the highest percentage answer). Where participants gave equal weighting to two or all of the strategies (n = 105) they were excluded from the following analysis.

Although a chi-square test showed no association between the most important health aspect used and the main strategy used when making the SRH judgement, z-tests (Bonferroni corrected for multiple comparisons) showed significant differences in the proportion of participants using comparison strategies when two specific health aspects

were used to make the SRH judgement. When using engagement with healthy behaviours to judge their health, a significantly greater proportion of participants (6.8%) used comparison to others as their main strategy for making the SRH judgement than used comparison to previous self (1.4%). Of the 19 participants who used engagement with healthy behaviours to judge their health, 14 did so by comparing to others, 3 did so by comparing to themselves previously and 2 did so using some other strategy. In the opposite direction, when using pain to judge their health, a significantly greater proportion of participants (6.3%) used comparison to previous self as their main strategy for making the SRH judgement than used comparison to others (1.4%). Of the 23 participants who used pain to judge their health, 14 did so by comparing to others, 3 did so by comparing to themselves previously and 6 did so using some other strategy.

Discussion

Social Comparison Effects

This study primarily investigated the extent to which people make judgements about their general health through comparing their health to that of other people and the exact cognitive mechanisms used in this comparison process. The vast majority of participants (84%) said that they compared to others (mainly friends, family, and people around their age) to some extent when answering a self-rated health (SRH) question. On average, participants who compared to others said that the comparison contributed 44% to the judgement. Regression analyses indicated that social comparisons accounted for 3-6% of the variance in SRH judgements but were inconclusive as to how people compare to others when making this judgement. Beliefs about the health of others in comparison groups that were shared across participants varied considerably. The correlation between objective health and SRH was significantly stronger for participants who stated that they did not compare to others than for those that did suggesting that SRH is a better proxy for actual health when people do not use social comparison to make the SRH judgement.

Main Implications

Taken together, these findings have a number of implications regarding our understanding of how people make inaccurate judgements about their health, the measurement of SRH, and future investigations of comparison effects. As discussed in the introduction of this chapter, previous research has shown that people often over- or underestimate their actual health when answering SRH questions. This finding has previously been explained as being the result of people comparing to others when making the SRH judgement and using different comparison groups when doing so. The findings reported here would support this hypothesis: people tend to compare to people that are around them – friends, family and people their age – and so the people that make up the comparison group (and their health) will be different for everyone.

However, previous research has never fully explained precisely why the use of different comparison groups may lead to over and under-estimations of health. It is proposed here that over or under-estimation of health occurs when the health of people in the comparison group is not representative of the health of people in the general population. Therefore, it is not so much that people make inaccurate judgements about their health because they use different comparison groups; it is more about how representative these comparison groups are. For example, the distribution of health in one person's comparison group may be very similar to the distribution of health in the general population and so that person may make an accurate judgement about their own health when comparing it to the health of the people in their sample. However, another person may have a comparison group comprised of people who are very ill and who are therefore not representative of the whole distribution of health in the general population but rather just the extreme tail of it. If that person is also in poor health but is not as bad as their comparison sample they may judge that they are much better off than they actually are, regardless of how they make this comparison (i.e., whether they use rank or average-based

strategies). This person will have come to an overestimation of their health because they have based the estimation on how their health compares to others who are not representative of the general population. Underestimation can occur similarly when a person who is in good health compares to others who are in exceptionally good health because, compared to these others, their health status ranks quite low/they are far below average but compared to the general population their health status actually ranks quite high/is above average.

One may think then that a way to reduce inaccurate health judgements would be to ask people to compare to the general population (or possibly people their age in the general population as health and age are strongly related) when making SRH judgements so that they do not use non-representative comparison samples. This is certainly what some surveys have done. For example, the British Household Panel Survey asks "compared to people of your own age, would you say that your health has on the whole been: excellent, good, fair, poor or very poor?". However, the rationale behind this question wording comes from the research suggesting that it is simply the use of different comparison samples that leads to inaccurate judgements. Therefore, it is assumed that people will make more accurate judgements about the actual state of their health by asking them to compare to a comparison group that is shared amongst all respondents such as the general population or shared amongst groups of respondents such as people of the same age (and not, for example, friends and family which differs from person to person). However, just as the distributions of health in people's individually constructed comparison samples can vary and either be representative of the actual distribution of health in the general population or not, so can people's beliefs about this actual distribution of health, possibly because they are using their individual samples to generalise. Figure 5.1 shows exactly this - participants' beliefs about the average health of both the general population and people their age varied widely meaning that many participants had inaccurate perceptions about the actual

distribution of health in these groups. It is therefore possible that people will still come to inaccurate judgements about their health status when they are asked specifically to compare to a more global sample, such as the general population or people their own age, because their beliefs about the distribution of health in these groups are incorrect. For example, if two people are in identical health to each other but one believes that 20% of people their age are in better health than them whilst the other believes that 80% are in better health they are likely to come to completely different subjective health judgements. Whilst one of them may be right in their estimation and accurately report their health, the other may not be and this could lead to over- or underestimating. It is therefore not enough to ask people to compare to a comparison group that is shared amongst all respondents when making the SRH judgement. Surveys also need to ask respondents about their beliefs about the distribution of health in this comparison group so that their answers may be adjusted for varying beliefs. Alternatively, information about the actual distribution of health in the comparison group could be provided alongside the SRH question. This could be tested in future research to see whether controlling for both comparison group and varying beliefs about others produces SRH judgements that more accurately reflect objective health.

The current findings also have implications for researchers investigating comparison effects using large datasets. Firstly, it should not be assumed that people compare their health to the average health of others in the comparison group as has been done previously because it is possible that people base their SRH judgement on their ranked position within the comparison group. Secondly, although data from large-scale studies can be used to construct comparison groups, the distribution of health in these groups may vary considerably from people's believed distribution, which is the one they would use to base the judgement about their own health on. Therefore, studies may be missing comparison effects or underestimating their strength through using this methodology.

This study also investigated group differences in comparison groups used when making the SRH judgement, the extent to which people compare their current health to their previous health when making SRH judgements and the aspects of health that people use to both make the SRH judgement and when making comparisons. These findings will be discussed next.

Group Differences in Chosen Comparison Group

There were significant differences across some of the demographic groups in the use of three of the most common comparison groups used when making the SRH judgement: family, people around the same age and the general population. A greater proportion of participants under 30 chose family as their most important comparison group than of participants over 40, whereas a greater proportion of participants over 40 chose people their age as their most important comparison group than of participants under 30. A greater proportion of healthy participants chose family as their most important comparison group than of participants in poor health whereas a greater proportion of participants in poor health chose people their age and the general population as their most important comparison group than of healthy participants. A greater proportion of White participants also chose the general population as their most important comparison group than of non-White participants.

Group differences in the choice of comparison group have largely been unexplored in the previous literature with the exception of Kaplan and Baron-Epel (2003) who found that participants in good health tended to compare their health to friends whilst those in poor health compared to people their age as is found here. Findings from the current study confirm that people use different comparison groups when evaluating their health and that there are systematic differences in the choice of comparison group.

Comparisons to Previous Health

The majority of participants (90%) said that they compared their current health to their previous health to some extent when answering the SRH question. On average, participants who compared to their previous health said that the comparison contributed 47% to the judgement. Women and participants over 40 based a higher percentage of their SRH judgement on how their current health compared to their previous health than men and participants under 30 respectively. These findings provide evidence of further context effects that may affect SRH judgements.

Health Aspects Used in Judgement

General physical condition was the health aspect that was used the most when making SRH judgements and comparing current health to both others and previous health. There were significant differences across some of the demographic groups in the use of the three most common health aspects when making the SRH judgement: general physical condition, the presence of health problems and the absence of health problems. A greater proportion of university educated participants chose the absence of health problems as the most important aspect used when making the SRH judgement than of non-university educated participants. A greater proportion of non-White participants chose the presence of health problems as their most important aspect than of White participants (also found by Krause & Jay, 1994). A greater proportion of healthy participants chose general physical condition and the absence of health problems as their most important aspect than of participants in poor health. The data suggested that the health aspects that participants most commonly compared to others and to their previous health when judging their general health were very similar. In most cases, participants were no more likely to use one comparison type (others vs previous health) over the other when using a specific health aspect to make the SRH judgement. However, participants who used pain to judge their health were more likely to do so by comparing their current experience of pain to their

previous experiences than by comparing their experience to others. On the other hand, participants who used engagement with healthy behaviours to judge their health were more likely to do so by comparing to others than they were to their previous engagement with such behaviours. However, the number of participants using these two aspects to judge their health was very small (pain: *n* = 23, healthy behaviours: *n* = 19) and so these results should be interpreted with caution.

Study Limitations

There were some limitations of the current study. Firstly, quite a basic measure of objective health was used (the number of symptoms recently experienced and the number of chronic conditions experienced) which only encompassed one aspect of health – illness. This measure has been used in a number of previous studies to investigate the relationship between subjective and objective health and to explore comparison effects. However, other, possibly standardised measures (such as the SF-36 or the Health Utilities Index Mark 3, HUI: Feeny, Furlong, Boyle, & Torrance, 1995) that encompass other aspects of health such as physical functioning, general physical condition and engagement with healthy and unhealthy behaviours could have been used instead of or alongside the objective health measure used in this study. Secondly, because objective health and the average health of others were measured using different scales, the distance of each participant's health from their believed average health of others could not be calculated and used in the analyses investigating how people compare to others as in previous studies reported in this thesis. However, this was mainly an issue for calculating the correlations which were not the main analyses; the main findings from the regression models would not have been as affected as both objective health and the average health of others were included in the model.

Conclusions

To conclude, this study provides evidence that the majority of people compare to others when judging their health and base their judgement partly on this comparison. The

findings suggest that self-rated health may not approximate actual health well when judged through a comparison to others because doing so may lead to over-and underestimations of health when the health of the comparison sample selected is skewed or beliefs about others are not accurate. As SRH is widely used in economics research, policy making and as an outcome measure in clinical trials, this could lead to a number of issues such as a reduced ability to compare health ratings across groups and to distinguish healthy from ill people. Comparison effects may be controlled for by both asking people to compare to a shared comparison group, such as the general population, and eliciting beliefs about the health of others so that health ratings can be adjusted in line with the same underlying scale. However, where there is room in surveys for longer health measures it is perhaps advisable to use more objective measures such as the number of symptoms and/or chronic conditions experienced (possibly also taking into account severity) or standardised measures such as the SF-36 or HUI.

Chapter 6: Social Support

Overview

The two studies reported in this Chapter were planned just as the other studies reported in this thesis were in that the main aim of these studies was to explore whether participants compared to others when making specific judgements. However, other findings that did not relate to the core aims of the study but which had important implications for the measurement of received support were found during data analysis. As comparison effects were not found in either study, the data was used to publish the other, unrelated findings instead. This overview gives a detailed account of the original rationale behind the studies, the exploration of comparison effects that was undertaken, and the new unrelated findings that were uncovered. This is followed by the published manuscript of the paper explaining the new findings.

Social support has been studied intensively across disciplines for over fifty years. Despite this, little is known about the cognitive mechanisms underlying social support processes and in particular their link to health (Sarason, Sarason, & Gurung, 2001; Thoits, 2011). Studies have consistently shown weak to moderate correlations between the amount of support a person receives and their perceptions of this support - specifically their beliefs about the availability of support and their satisfaction with support (for a review see Haber, Cohen, Lucas, & Baltes, 2007). Furthermore, although there is a strong and well documented association between social support and both physical and mental health, research has shown that only support perceptions (and not the amount of support received) are consistently associated with positive health outcomes (e.g., Barrera, 2000; Holt-Lunstad, Smith, & Layton, 2010; Uchino, 2004; 2009; Uchino, Bowen, Carlisle, & Birmingham, 2012).

Despite these findings, it seems intuitive that people's perceptions of the support that they receive should be strongly related to that support and that if perceived support is associated with health then received support should be too. It was hypothesised that stronger associations between received support and both perceived support and health existed but that the way previous studies had investigated them had led to the full extent of these relationships not being seen. In the majority of these studies, received support was measured simply as the number of supportive behaviours received. Factors that may affect the received-perceived support relationship and possibly also the relationship between received support and health (such as the need for support, how the amount of support received compares to the amount other people receive and the quality of the support) had not been taken into account previously. Therefore, the initial aim of the social support studies was to investigate whether the relationships between received support and both perceived support and health were stronger when these factors were taken into account, with the specific aim of investigating any social comparison effects. It was hypothesised that it is not simply the absolute amount of support received that may influence support perceptions and health but how this amount compares to the amount of support that other people are believed to receive. If true, this could, for example, explain why the correlation between received and perceived support is weaker than expected as an individual may receive a lot of support but they could still not be satisfied with it because they think that other people receive more than them. Similarly, an individual may receive little support but still be satisfied with it because they think that they receive more than others.

In Study 1, participants were asked how many times in a typical month they received two types of support (absolute received support measure) and answered questions eliciting their beliefs about the distribution of how frequently people in the general population receive the support types. Each participant's rank amongst their believed distribution and how much the amount of support they received differed from the mean of this distribution

was then calculated using their answers to these questions (this distribution elicitation methodology is explained in detail in Chapter 2). Participants also completed five perceived support measures and one measure of their physical and mental health. Regression analysis was carried out as in previous chapters: models including age, gender, absolute and rank and age, gender, absolute and distance from the mean variables predicting the seven outcome variables were calculated. Unfortunately the vast majority of the models (76%: 22 out of 28) were not significant, i.e., the regression model including the independent variables did not predict the outcomes any better than the baseline model. Of the six significant models, the only significant predictor was the absolute amount of support received but generally the variance accounted for by these models was low (R^2 = .048-.067). Overall, the results replicated previous findings but provided no support for the hypotheses being tested.

However, when planning Study 1 it seemed intuitive that receiving support when it was not needed could be unbeneficial and have negative rather than positive consequences. Along with the number of times participants received support they were also asked on how many times they had actually needed it and those who stated that they did not need support were excluded from the analysis. However, this did not control for participants who needed support but received more support than they needed whom, under the above assumption, may also find extra, unneeded support unbeneficial. The data from Study 1 was re-analysed to investigate whether taking the need for support into account and also excluding participants who received more support than they needed, had any effect on the correlations between received and perceived support. A measure of the proportion of times support was received when needed was calculated by dividing the number of times participants stated they received support by the number of times they stated that it was needed. Participants who received more support than they needed, i.e., those with proportion values over 1 were then excluded. On average, the correlations

between received and perceived support doubled when the proportion measure was used as the measure of received support compared to when received support was measured solely as the number of supportive behaviours received. If receiving more support than is needed is associated with negative support perceptions, then received and perceived support will only be positively correlated up to the point where support needs are met. Therefore, if the received-perceived support relationship is affected by the need for support, as is suggested here, then this relationship will be quadratic and not linear. This can explain why previous studies investigating the received-perceived support relationship have found such low correlations. These studies are likely to include people whose support needs are met, people who receive less support than they need (under-supply) and people who receive more support than they need (over-supply). However, as previously stated, the majority of these studies used measures of received support that only measure the amount of support received and do not take the need for support into account. Without controlling for the need for support, the negative perceptions experienced by people receiving an over-supply of support will effectively cancel out the positive correlation between received and perceived support expected in the remainder of the participants.

It was not possible to investigate whether controlling for the need of support had any effect on the relationships between the social comparison variables and perceived support and health using the data from Study 1. This was because, unlike received support where participants had been asked about both the number of times support was needed and received, the distribution elicitation questions simply asked about the number of times different percentiles of the population received support and not received support when they needed it. The decision was made to run a second study using questions that were more tailored to measuring support received when needed to see whether 1) controlling for participants' support needs had any effect on the relationships between the social comparison variables and perceived support and health and 2) whether the increase in the

strength of the received-perceived relationship resulting from controlling for participants' support needs could be replicated.

Unfortunately, the absolute, rank, and distance from the mean variables for both support types were so highly correlated in Study 2 that it was not possible to carry out the regression analyses to investigate the hypotheses concerning social comparison effects. However, as in Study 1, the results showed that the strength of the received-perceived support relationship was much stronger when the need for support was controlled. As this was a novel finding, it was written as a separate paper (published in *Personality and Individual Differences*) which is printed below. This gives full details of the two studies undertaken. Table numbers have been changed to align with the other table numbers in this thesis.

When is Received Social Support Related to Perceived Support and Well-Being? When it is Needed

Abstract

How do perceptions of being supported relate to the amount of social support received? Received and perceived support have generally been found to be only moderately related. Previous research has however focused on the amount of support received regardless of whether it was needed. We hypothesized that a measure of support received when needed would predict perceived support and well-being better than would an unqualified measure of received support. Study 1 found that correlations between received support and perceived support measures were, on average, twice as high when received support was measured as the proportion of times support was received when needed (average *r*=.54) than when it was measured as the number of times support was received (average *r*=.28). Similar results were found for correlations between received support and mental health

which rose from *r*=.04 to *r*=.31 when need for support was considered. Study 2 replicated the strong relationship between support received when needed and both perceived support and mental health. Received support measures should be adapted to take the need for support into consideration in future investigation of these relationships. Social support interventions may only be beneficial if the recipient's support needs are not already being met.

Introduction

This paper is concerned with the relationship between received and perceived support and with the relationship between both received and perceived support and well-being. We operationalise received and perceived support as they are most commonly measured: received support as the quantity of supportive behaviours received by an individual (Haber, Cohen, Lucas, & Baltes, 2007) and perceived support as both the satisfaction with support and the availability of it (Sarason, Sarason, & Pierce, 1990).

There has been much interest in the relationship between received and perceived support because of the strong and well documented link between social support and health; people who are more socially integrated tend to be healthier, both physically and mentally, than those who are more socially isolated (Barrera, 1986; House, Landis, & Umberson, 1988; Uchino, 2009). A number of theories have been put forward to account for this, the most dominant being stress buffering theory (cf. Barrera, 1986; Cohen & Wills, 1985; Cutrona & Russell, 1990; Thoits, 1986). The theory proposes that social support acts as a buffer that protects people against the physical and mental effects of stress caused from experiences such as illness or other life events. It suggests that the relationship between received and perceived support should be relatively strong and that both positive perceptions of support and receipt of support should lead to stress-buffering effects (Haber et al., 2007, Lakey & Cohen, 2000). However, the relationship between received and perceived support, although significant, has been consistently found to be relatively mild.

For example, a meta-analysis of 23 studies found the average correlation between perceived and received support to be $r = .35$, $p < .001$ (Haber et al., 2007). Furthermore, whereas perceived support is consistently associated with positive health outcomes (e.g., Barrera, 2000; Holt-Lunstad, Smith, & Layton, 2010; Uchino, 2004; 2009; Uchino, Bowen, Carlisle, & Birmingham, 2012), the relationship between received support and health has been shown to be very inconsistent with nonsignificant and even negative associations often being found (Bolger & Amarel, 2007; Uchino, 2009). It is therefore unsurprising that interventions that have been developed based on this theory, under the assumption that increasing received support will lead to better health, have provided mixed results (Barrera, Glasgow, McKay, Boles, & Feil, 2002).

Many studies and evaluations of the relationships between received support, perceived support and health use measures of received support such as the Inventory of Socially Supportive Behaviors (ISSB; Barrera, Sandler, & Ramsey, 1981), which is the most widely used and well-validated measure of received support (Haber et al., 2007; Gottlieb & Bergen, 2010), but which only measures the amount of support received. This has meant that factors relating to the support received that may affect this relationship (such as whether it was needed or the quality of it) have been largely unexplored. Here we propose and test the hypothesis that the relationship between received and perceived support is affected by the need for support. Specifically, we hypothesize that people's perceptions of the support they receive are based not on the number of times they receive support but on the number of times they have received it relative to the number of times they have needed it. We also hypothesize that received and perceived support will be positively correlated only up until the point where support needs are met. Beyond this (i.e., when people experience an oversupply of support), we expect that the relationship may break down (i.e., become absent or even negative) and we therefore propose that previous tests

of the strength of the received-perceived support relationship may be inaccurate if the need for support has not been controlled for.

Although these hypotheses appear not to have been tested in the context of the received-perceived support relationship, there is some evidence to suggest that the relationship between received support and health may be stronger when the need for support is taken into account for the reasons outlined above. Studies have shown that both an under- (Jou & Fukada, 2002) and over-supply (Reynolds & Perrin, 2004) of support can lead to negative health outcomes. Therefore, analyses that fail to take the need for support into account may produce weaker correlations between received support and health because any positive effects of receiving additional support when it is needed may be counteracted by reduced, absent or even negative effects of receiving support when it is not. Wolff, Schmiedek, Brose and Lindenberger (2013) found support for this hypothesis and demonstrated how taking the need for support into account improved the strength of the relationship between received support and health. They found no significant relationship between the amount of support received and either physical health or emotional well-being, but obtained a significant, quadratic relationship between these two outcomes and the balance of received and needed support (i.e., the difference in the number of times support is needed and actually received).

The primary aim of the current study is therefore to investigate whether the received-perceived support relationship is stronger when the need for support is taken into account as appears to be the case with regard to the relationship between received support and health. It is possible that previous findings of weak relationships between received support and both perceived support and health have been due to a common cause – namely the way received support has been measured. We also aim to provide more evidence that the relationship between received support and health is also stronger when the need for support is considered.

Study 1

In Study 1 we directly compare the relationships between received support, support received when needed, perceived support and health. Based on previous findings we hypothesized that there would be a significant but mild correlation between received and perceived support and that this relationship would strengthen when the need for support is taken into account. Due to previous inconsistent findings we were unsure as to whether or not a significant correlation between received support and health would be found but we expected a significant, positive correlation between these constructs when the need for support is taken into account. We further investigated differences in the strength of the relationships between received support, support received when needed, perceived support and health using regression analyses. This allowed for us to determine whether our measure of support received when needed (described below) predicted perceived support and health outcomes over and above received support alone. We hypothesized that our measure of support received when needed would be a much stronger predictor of perceived support and health outcome measures than received support.

Method

Participants: The 198 participants had a mean age of 32.4 years (SD = 12.8, range: 18-65 years), were predominantly White (76%) and 47.5% were male. Participants were mainly college educated (69%), were all resident in the U.S., and completed the study from 41 different states. Sample size calculations were based on detecting the weakest effect, i.e., the correlation between received and perceived support, which a meta-analysis identified to be r = .35 on average (Haber et al., 2007). Calculations showed that at least 121 participants would be needed to have a 99% chance of detecting a correlation of .35 and that for multiple regressions with 4 predictor variables at least 174 participants would be needed to have a 99% chance of detecting a medium sized effect (a correlation of .35 indicates a medium sized effect; Cohen, 1988).

Procedure: Participants were recruited online through Amazon Mechanical Turk (www.mturk.com) – an online crowdsourcing platform where "workers" choose tasks to complete in exchange for money or Amazon vouchers. Mechanical Turk workers have been shown to produce high quality data in psychological experiments (Buhrmester, Kwany & Gosling, 2011) and to be more representative of the U.S. population than university undergraduates typically used in psychological research as well as other internet samples in general (Paolacci, Chandler, & Ipeirotis, 2010). Mechanical Turk has also been found to be a reliable source of experimental data specifically in the area of judgment and decision making (Paolacci et al., 2010). Participants were asked to complete an online questionnaire that comprised questions on needed support, received support, perceived support, mental and physical health and demographic questions (age, gender, level of education and ethnicity). They received $1.00 on completion of the study which took 15-20 minutes to complete; this payment was in line with typical Mechanical Turk payments. As no standardised measures of support received when needed could be found, two specific supportive behaviours (having someone listen to you talk about your private feelings and having someone pitch in to help you do something) each representing a different type of support (emotional support and tangible assistance respectively) were chosen from the ISSB as the focus of the study. Participants were asked the following:

Needed and received support: Participants were asked the following questions about their need and receipt of emotional support: "In a typical month, how many times [do you need]/[does] someone to listen to you talk about your private feelings?", and tangible support: "In a typical month, how many times [do you need]/[does] someone to pitch in to help you do something that needs to be done?".

Perceived support: As we had asked participants about their receipt of two specific supportive behaviours we also asked about their perceptions relating specifically to these behaviours. We asked participants to rate on six-point scales how satisfied they were with

the support they received from people who listen them talk about their private feelings and from people who pitch in to help them do something that needs to be done right away, how available these people were to them and how satisfied they were with the availability of these people. Perceived support was also measured using two standardised but non-support type specific scales; the Multidimensional Scale of Perceived Social Support (MSPSS; Zimet, Dahlem, Zimet, & Farley, 1988, see Appendix 3) and the satisfaction subscale of the Arizona Social Support Interview Schedule (ASSIS-S; Barrera, 1981, see Appendix 4). The 12 item MSPSS measures the current availability of social support from family, friends and significant others rated on a seven point scale from (1) very strongly disagree to (7) very strongly agree. High total scores on this measure indicate high levels of perceived social support. The MSPSS has been found to have strong internal (multiple tests show Cronbach's alpha values ranging from .84 to .92; Zimet et al., 1988, Zimet, Powell, Farley, Werkman, & Berkoff, 1990) and test-retest reliability (Cronbach's $\alpha = .85$; Zimet et al., 1988) and moderate construct validity (demonstrated by a significant, negative relationship between the scale and depression symptoms: $r = -.25$, $p < .01$; Zimet et al., 1988). The ASSIS-S is a six item measure of satisfaction with support received in given situations during the past month rated on seven point scale from (1) very dissatisfied to (7) very satisfied. High total scores on this measure also reflect high levels of perceived social support. This satisfaction subscale has been shown to have moderate test-retest reliability ($r = .69$, $p < .001$; Barrera, 1981).

Health: Health was measured using the Short Form-36v2 Health Survey (SF-36v2; Ware, Kosinski, Bjorner, Turner-Bowker, Gandek, & Maruish, 2008, see Appendix 5) which is a generic measure of both physical and mental health status and the most widely used health survey in the world (Ware et al., 2008). The 36 items evaluate four areas of physical health: physical functioning, role-physical, bodily pain and general health and four areas of mental health: vitality, social functioning, role-emotional and mental health. Test scoring

produces separate scores for physical (physical component summary: PCS) and mental health (mental component summary: MCS) and scores range from 0 ("poorest" health status) to 100 ("best" health status). The SF-36 Health Survey has had extensive reliability and validity testing (for a comprehensive review see Ware & Gandek, 1998). Reliability estimates (Cronbach's alpha) for the physical and mental component summaries usually exceed .90 (Ware, Kosinski, & Keller, 1994); exact numbers given in the user guide are .95 for the PCS and .93 for the MCS (Ware et al., 2008).

Statistical analysis: We first predict that using a measure of received support that takes the need for support into account will produce stronger correlations with perceived support and health and will predict these outcomes over and above traditional measures of received support. We calculated such a measure – the proportion of times support is received when needed – by dividing the number of times participants stated they received support by the number of times they stated that it was needed. We used a relative measure as opposed to an absolute one as we hypothesise that people receiving support, for example, 50% of the time they need it will have the same perceptions of this support regardless of the absolute numbers of times they receive and need support. Use of this measure meant that participants who stated that they didn't typically need support ($n_{emotional}$ = 14, $n_{tangible}$ = 13) needed to be excluded from the study as accurate proportions could not be calculated for these participants. Secondly, we predict a positive, linear relationship between support received when needed and both perceived support and health such that perceptions of support become more positive and wellness increases as the proportion of times support needs are met increases to 1 (i.e., support received = support needed). Due to the use of multiple analyses a conservative alpha level of .01 was used for all statistical tests.

Results and Brief Discussion

Descriptive Statistics: For ease of comparison we provide a summary of the means and standard deviations of participants' responses to variables included in both Study 1 and Study 2 in Table 6.1 below.

Table 6.1.
Means and Standard Deviations (in parentheses) of Responses to All Social Support and Health Variables in Study 1 and Study 2.

	Study 1		Study 2	
	Emotional	Tangible	Emotional	Tangible
Received Support	6.12 (8.04)	5.71 (7.38)	31.39 (15.77)	28.76 (15.23)
Satisfaction	4.60 (1.14)	4.38 (1.18)	4.46 (1.21)	4.29 (1.24)
Availability	4.32 (1.20)	4.09 (1.15)	4.28 (1.06)	3.97 (1.06)
Satisfaction with Availability	4.42 (1.35)	4.15 (1.35)	4.32 (1.30)	4.12 (1.33)
MSPSS Average Item Score	5.08 (1.22)		5.08 (1.22)	
ASSIS-S Scores	31.37 (6.62)		30.98 (6.33)	
Physical Health	54.40 (7.57)		53.71 (8.90)	
Mental Health	43.41 (11.36)		45.65 (11.03)	
n	198		202	

Note. In Study 1 received support was measured as the number of times in a typical month support is received but in Study 2 it was measured as the number of times support was received on the last 50 occasions it was needed. Support satisfaction, availability and satisfaction with availability were all measured on six-point scales where high scores represent high levels of these constructs. MSPSS = Multidimensional Scale of Perceived Social Support, ASSIS-S = Arizona Social Support Inventory Schedule – Satisfaction subscale, physical health is measured by the SF-36v2 Health Survey Physical Component Summary and mental health is measured by the SF-36v2 Health Survey Mental Component Summary. For perceived support scales higher scores represent higher levels of support and for health scales higher scores represent better health.

It is not possible to compare the amount of support received by participants in Study 1 and 2 directly, as support was measured over different periods (see Study 2 below). In Study 1 we asked participants about the number of times they received support in a typical month and in Study 2 we asked about the number of times they had received support on the last 50 occasions when they had needed it. It is somewhat difficult to ascertain whether the amount of support received by our internet samples is typical of the amount of support received by the general population. This is because the scale that we took our received support questions from (the ISSB) asks respondents to rate the frequency with which they have received the support type in the last four weeks on a five point scale (1 = "not at all", 2 = "once or twice", 3 = "about once a week", 4 = "several times a week", 5 = "about every

day") rather than specifically asking for the number of times support is received as we have done in Study 1. In their paper discussing the development of the ISSB, Barrera et al. (1981) state that the average rating given to the emotional support item used in the present study was 2.72 (SD = 1.23, n = 71) and to the tangible support item was 2.32 (SD = 0.97, n = 71). This suggests that their participants received these two types of support somewhere between 1 and 4 times in the last 4 weeks which is slightly lower than our participants' receipt of this support ($M_{emotional}$ = 6.12, $M_{tangible}$ = 5.71).

There were only small differences across studies in the average responses given on the five perceived support measures. As we created three of the measures ourselves (the satisfaction, availability and satisfaction with availability questions) we are unable to compare these responses to those from a general population sample. However, during development of the MSPSS, Zimet et al. (1988) found the average item score of their sample (n = 275) to be 5.60 (SD = 0.86) which is very similar to ours ($M_{emotional}$ = 5.08, $M_{tangible}$ = 5.08). Average scores on the ASSIS-S were not reported in papers describing the development of the scale (Barrera, 1980; Sandler & Barrera, 1984) and so it is not possible to compare scores on this measure across populations.

There was also very little difference in the physical and mental health of participants across our two studies. The mental and physical component summaries of the SF-36v2 were each designed to have a mean of 50 and a standard deviation of 10 (Ware & Gandek, 1998); our participants were slightly better than average in terms of their physical health (M_{Study1} = 54.40, M_{Study2} = 53.71) but were poorer than average in their mental health (M_{Study1} = 43.41, M_{Study2} = 45.65).

Correlations: Correlations (Spearman for ordinal outcome variables and Pearson for continuous) were undertaken between all variables for each type of support (see Table 6.2).

Table 6.2.

Correlations between Received Support, Proportion of Times Support is Received when Needed and All Outcome Variables for Each Support Type.

	Emotional Support				Tangible Support			
	Support Received (all)	Support Received (after exclusion)	Support Proportion (all)	Support Proportion (after exclusion)	Support Received (all)	Support Received (after exclusion)	Support Proportion (all)	Support Proportion (after exclusion)
Satisfaction	.25*	.33**	.29**	.48**	.21*	.20*	.45**	.49**
Availability	.42**	.48**	.48**	.60**	.29**	.28**	.47**	.53**
Satisfaction with Availability	.35**	.43**	.45**	.56**	.27*	.28**	.49**	.59**
MSPSS Scores	.15	.29**	.28**	.54**	.17	.12	.21*	.50**
ASSIS-S Scores	.13	.24*	.20*	.56**	.22*	.16	.18	.50**
Physical Health	-.07	.05	.00	.07	.02	.00	-.11	.02
Mental Health	.00	.04	.18	.36**	.08	.03	.08	.26*
n	184	154	184	154	185	168	185	168

*Note. MSPSS = Multidimensional Scale of Perceived Social Support, ASSIS-S = Arizona Social Support Inventory Schedule – Satisfaction subscale, physical health is measured by the SF-36v2 Health Survey Physical Component Summary and mental health is measured by the SF-36v2 Health Survey Mental Component Summary. For perceived support scales higher scores represent higher levels of support and for health scales higher scores represent better health. *p<.01, **p<.001 (two-tailed).*

***Perceived support**:* Correlations using all our participants (except those who stated that they do not need support) showed that, as hypothesized, both received emotional and tangible support correlate weakly with our five measures of perceived support (average: emotional: r = .26, tangible: r = .23). All correlations were significant apart from the correlation between received emotional support and ASSIS-S scores and the correlations between both types of received support and MSPSS scores. These findings follow the pattern of results previously seen in the literature investigating the received-perceived support relationship. As predicted, these correlations improved when need for support was taken into consideration. The average correlation between the proportion of times support is received when needed and our perceived support measures was r = .34 for emotional support and r = .36 for tangible support. Although improved, these correlations are still weak. We hypothesize that this is because received and perceived support are strongly and positively correlated up until the point where support needs are met. Beyond this (i.e., when people are over-supplied with support) we hypothesise that the received-perceived support relationship breaks down in some way; it perhaps becomes weaker, absent or even negative. We investigate this by removing the participants who stated that they receive more support than they need ($n_{emotional}$ = 30, $n_{tangible}$ = 17) from the analysis. When we do this the correlations between received support and perceived support measures remain largely unchanged for tangible support (average: r = .21) and increase only slightly for emotional support (average: r = .35). However, the correlations between the proportion of times support is received when needed and our measures of perceived support increase considerably (average: emotional: r = .55, tangible: r = .52). Taken together these findings suggest that the strength of the relationship between received and perceived support is affected by whether or not the support received is needed.

Health: Contrary to our perceived support findings we found no significant correlations between scores on the physical component summary of the SF36v2 health

survey and either received support or the proportion of times support is received when needed. Excluding participants who received an over-supply of support did not change these findings. These results are in line with other studies that have shown no association between received support and physical health but contradict findings of Wolff et al. (2013) who showed a significant relationship between physical health complaints and the balance of needed and received support. We suspect that these findings may be due to the different way physical health was measured in the two studies and this possibility is discussed in the general discussion.

The correlations between scores on the mental component summary of the SF36v2 health survey and both received support and the proportion of times support is received when needed were also not significant. When participants who received an oversupply of support were excluded there was no difference in the correlations between received support and mental health scores. However, the correlations between the proportion of times support is received when needed and mental health scores increased substantially following this exclusion (for emotional support, from r = .18 to r = .36 and tangible, from r = .08 to r = .26), supporting our hypotheses and in line with Wolff et al.'s (2013) findings.

Regressions: Excluding participants who received more support than they needed, we used multiple linear regression (using the enter method) to predict all seven outcomes from age, gender, received support and the proportion of times support is received when needed. The number of remaining participants exceeded the minimum number required to have a 95% chance of detecting an effect (n = 129). The primary aim of these analyses was to investigate whether the proportion of times support is received when needed predicted the outcomes over and above support received. We also control for age and gender by adding these variables to the model as it is possible that they could be factors that moderate the relationships between received support and both perceived support and health (Haber et al., 2007; Wolff et al., 2013).

For the three ordinal outcome variables (satisfaction with support, availability of support and satisfaction with availability of support) we also conducted ordinal regression analyses (using the polytomous universal model and logit function). The findings were the same as the linear models in terms of whether the overall model and predictors in the model were significant and so, for ease of interpretation, we only report the linear regression analyses here. Table 6.3 shows 14 two-step models, one for each outcome and support type. In the first step age, gender and received support were included in the model and in the second step the proportion of times support is received when needed was added to the model. For brevity we do not include the constant, age or gender coefficients in the table. Neither age nor gender were significant predictors of our outcomes in any of the models or steps. The constant was significant in all models and steps. Tolerance values for all independent variables were above 0.10 suggesting that collinearity is not problematic.

Perceived support: For emotional support we see a general trend such that the variables entered into the model at Step 1 significantly predict the perceived support outcomes over and above the baseline model and that received support is the only significant predictor in the model (all outcomes apart from satisfaction with emotional support where there are no significant predictors). These models account, on average, for 9% of the variance (range: 5-13%). When the proportion of times support is received when needed is added to the model in Step 2, the effect of received support is removed, the variance accounted for by the model increases, on average, by 25% (range: 19-30%) and proportion becomes the sole predictor of the perceived support outcome measure. For tangible support all of the five models predicting perceived support outcomes only reach significance when proportion is added to the model in Step 2. Proportion is the sole significant predictor in all of these models which account for, on average, 32% of the variance (range: 26-42%). These results provide evidence of a much stronger relationship

between the proportion of times support is received when needed and perceived support than the number of times it is received.

Health: For both emotional and tangible support, none of the variables entered into the model at either step significantly predicted physical health scores but the proportion of times support is received when needed significantly predicted mental health scores. These models accounted for 13% (emotional support) and 8% (tangible support) of the variance.

Table 6.3.

Multiple Linear Regression Analyses Predicting Satisfaction with Support, Availability of Support, Satisfaction with Availability of Support, Scores on the Multidimensional Scale of Perceived Social Support (MSPSS), Scores on the Arizona Social Support Interview Schedule – Satisfaction Subscale (ASSIS-S) and Scores on the SF-36v2 Health Survey Physical and Mental Component Summaries (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Emotional Support				*Tangible Support*			
Predictors	ΔR^2	B (95% CI)	β	*p*	ΔR^2	B (95% CI)	β	*p*
Satisfaction								
Step 1	.05			.041	.03			.174
Received Support		0.04 (-0.01-0.10)	.21	.060		0.02 (0.00-0.05)	.11	.060
Step 2	.19**			**<.001**	.27**			**<.001**
Received Support		0.01 (-0.04-0.06)	.05	.707		0.00 (-0.02-0.02)	.01	.881
Proportion of Support		**1.67 (1.02-2.32)**	**.47**	**.001**		**2.08 (1.42-2.77)**	**.53**	**.001**
Availability								
Step 1	.13**			**<.001**	.03			.135
Received Support		**0.08 (0.05-0.12)**	**.35**	**<.001**		0.02 (0.00-0.07)	.13	.112
Step 2	.30**			**<.001**	.32**			**<.001**
Received Support		0.03 (0.01-0.07)	.15	.033		0.00 (-0.02-0.04)	.02	.833
Proportion of Support		**2.14 (1.53-2.69)**	**.58**	**.001**		**2.18 (1.69-2.67)**	**.58**	**.001**
Satisfaction with Availability								
Step 1	.12**			**<.001**	.03			.160
Received Support		**0.09 (0.05-0.13)**	**.34**	**<.001**		0.03 (0.00-0.09)	.14	.137
Step 2	.28**			**<.001**	.39**			**<.001**
Received Support		0.04 (0.00-0.07)	.14	.028		0.00 (-0.02-0.04)	.01	.878
Proportion of Support		**2.45 (1.82-3.13)**	**.57**	**.001**		**2.89 (2.29-3.45)**	**.64**	**.001**
MSPSS								
Step 1	.10*			**.001**	0.03			.210
Received Support		**0.74 (0.33-1.19)**	**.28**	**<.001**		0.25 (-0.05-0.59)	.12	.060
Step 2	.23**			**<.001**	.24**			**<.001**
Received Support		0.27 (-0.11-0.65)	.10	.152		0.04 (-0.31-0.32)	.02	.774
Proportion of Support		**22.37 (15.16-29.96)**	**.51**	**.001**		**23.82 (16.34-31.64)**	**.50**	**.001**

Table 6.3. Continued.

ASSIS-S								
Step 1	.06			.027	.03			.218
Received Support		**0.30 (0.09-0.55)**	**.24**	**.006**		0.16 (0.03-0.34)	.16	.019
Step 2	.26**			**<.001**	.23**			**<.001**
Received Support		0.07 (-0.11-0.27)	.06	.429		0.06 (-0.07-0.18)	.06	.294
Proportion of Support		**11.32 (7.74-14.68)**	**.54**	**.001**		**10.78 (7.77-14.00)**	**.49**	**.001**
Physical Health								
Step 1	.02			.307	.03			.191
Received Support		0.09 (-0.11-0.28)	.06	.380		-0.01 (-0.15-0.13)	-.01	.890
Step 2	.00			.409	.00			.308
Received Support		0.06 (-0.18-0.30)	.04	.581		-0.01 (-0.16-0.12)	-.01	.835
Proportion of Support		1.24 (-2.99-5.42)	.05	.517		0.55 (-3.81-5.08)	.02	.800
Mental Health								
Step 1	.00			.904	.01			.650
Received Support		0.09 (-0.27-0.51)	.04	.632		0.04 (-0.23-0.32)	.03	.627
Step 2	.13**			**<.001**	.07*			.011
Received Support		-0.20 (-0.52-0.18)	-.09	.259		-0.05 (-0.33-0.22)	-.03	.622
Proportion of Support		**14.01 (7.58-21.20)**	**.39**	**.001**		**10.39 (4.79-15.90)**	**.26**	**.001**

Note. $n_{emotional}$ *= 154,* $n_{tangible}$ *= 168, CI = confidence interval, significant models and predictors are highlighted in bold.*p<.01, **p<.001*

Study 2

We aimed to replicate our findings of a strong relationship between the proportion of times support is received when needed and perceived support from Study 1 in Study 2. We also further investigated the relationship between support received when needed and health.

Method

Participants: The 202 participants had a mean age of 34.8 years (SD = 13.2, range: 18-73 years), were predominantly White (85%), and 47% were male. Participants were mainly college educated (75%), were all resident in the U.S., and completed the study from 40 different states. There was no significant difference in the average ages of participants in Study 1 (M = 32.37, SD = 12.81) and Study 2 (M = 34.77, SD = 13.22), $t(398) = -1.84$, $p = .066$. There was also no significant association between study and participants' gender ($\chi^2(1, N = 400) = 0.04$, $p = .920$) or between study and participants' education level ($\chi^2(2, N = 400) = 1.95$, $p = .577$). This suggests that participants from Study 1 and Study 2 did not differ in age, gender and education level.

Procedure: Participants were again recruited using Amazon Turk, using the same procedure and compensation as in Study 1. Participants followed the same procedure as in Study 1 apart from being asked the following questions instead of being asked how many occasions in a typical month they needed and received emotional and tangible support: "Consider 50 occasions that you have needed someone to listen to you talk about your private feelings. On how many of these occasions did someone provide this support?" and "consider 50 occasions that you have needed someone to pitch in to help you do something that needs to be done. On how many of these occasions did someone provide this support?". The questions were framed this way so that participants who do not need support on a monthly basis would not have to be excluded (as in Study 1) as we would have valid proportion measurements for all participants. Of course, it was not expected that

participants would be able to recall 50 specific occasions. Rather, our use of this wording was designed to encourage participants to respond in terms of the proportion of occasions on which they had received support when it was needed.

Results

All participants were included in the analysis as none stated that they received support on more occasions than they needed it (i.e., none answered the above questions with a figure over 50) as would be expected with how the question was worded. Again, an alpha level of .01 was used for all statistical tests.

Correlations: Correlations (Spearman for ordinal outcome variables and Pearson for continuous) were undertaken between all variables for each type of support. Table 6.4 shows a comparison of the correlations between the proportion of times support is received when needed and our perceived support and health outcomes from Studies 1 and 2. As can be seen, the correlations are very similar to, and replicate the findings from, Study 1.

Regressions: We carried out ordinal and multiple linear regression analyses as in Study 1 using age, gender and the proportion of times support is received when needed as predictors in the models. For brevity, we include the age and gender coefficients only for models where they are significant predictors. The constant was significant in all models. Tolerance values for all independent variables were above 0.10 suggesting that collinearity is not problematic.

Table 6.4.

Correlations between Proportion of Times Support is Received when Needed and All Outcome Variables for Each Support Type from Study 1 and 2.

	Satisfaction	Availability	Satisfaction with Availability	MSPSS Scores	ASSIS-S Scores	Physical Health	Mental Health
Emotional Support Proportion Study 1	.48**	.60**	.56**	.54**	.56**	.07	.36**
Emotional Support Proportion Study 2	.56**	.61**	.59**	.56**	.56**	-0.02	.37**
Tangible Support Proportion Study 1	.49**	.53**	.59**	.50**	.50**	.02	.26**
Tangible Support Proportion Study 2	.43**	.49**	.46**	.56**	.55**	-0.04	.37**

*Note. Study 1: $n_{emotional}$ = 154, $n_{tangible}$ = 168, Study 2: N = 202. MSPSS = Multidimensional Scale of Perceived Social Support, ASSIS-S = Arizona Social Support Inventory Schedule – Satisfaction subscale, physical health is measured by the SF-36v2 Health Survey Physical Component Summary and mental health is measured by the SF-36v2 Health Survey Mental Component Summary. For perceived support scales higher scores represent higher levels of support and for health scales higher scores represent better health. **p<.001 (two-tailed).*

Table 6.5.

Multiple Linear Regression Analyses Predicting Satisfaction with Support, Availability of Support, Satisfaction with Availability of Support, Scores on the Multidimensional Scale of Perceived Social Support (MSPSS), Scores on the Arizona Social Support Interview Schedule – Satisfaction Subscale (ASSIS-S) and Scores on the SF-36v2 Health Survey Physical and Mental Component Summaries (95% BCa Bootstrap Confidence Intervals and Standard Errors Based on 1000 Samples).

	Emotional Support				*Tangible Support*			
Predictors	ΔR^2	B (95% CI)	β	*p*	ΔR^2	B (95% CI)	β	*p*
Satisfaction								
Model	.28**			**< .001**	.22**			**< .001**
Proportion of Support		**1.95 (1.46-2.47)**	**.51**	**< .001**		**1.90 (1.25-2.51)**	**.47**	**< .001**
Availability								
Model	.34**			**< .001**	.25**			**< .001**
Proportion of Support		**1.96 (1.53-2.34)**	**.58**	**< .001**		**1.74 (1.17-2.30)**	**.50**	**< .001**
Satisfaction with Availability								
Model	.33**			**< .001**	.20**			**< .001**
Proportion of Support		**2.37 (1.81-2.93)**	**.57**	**< .001**		**1.98 (1.27-2.58)**	**.45**	**< .001**
MSPSS								
Model	.32**			**< .001**	.31**			**< .001**
Proportion of Support		**25.06 (18.38-31.79)**	**.55**	**< .001**		**25.80 (19.11-32.47)**	**.55**	**< .001**
ASSIS-S								
Model	.32**			**< .001**	.31**			**< .001**
Proportion of Support		**11.08 (8.51-13.55)**	**.55**	**< .001**		**11.33 (8.62-13.97)**	**.55**	**< .001**
Physical Health								
Model	.06*			.01	.06*			.01
Age		**-0.17 (-0.28--0.06)**	**-.25**	**< .001**		**-0.17 (-0.27--0.07)**	**-.25**	**< .001**
Gender		0.03 (-2.51-2.46)	.00	.98		-0.07 (-2.53-2.54)	.00	.94
Proportion of Support		-0.90 (-4.68-3.39)	-.03	.68		-1.64 (-5.80-2.42)	-.06	.48
Mental Health								
Model	.23**			**< .001**	.23**			**< .001**
Age		**0.22 (0.11-0.33)**	**.26**	**< .001**		**0.21 (0.11-0.31)**	**.25**	**< .001**
Gender		3.99 (1.33-6.75)	.18	.01		**4.50 (1.92-7.17)**	**.20**	**< .001**
Proportion of Support		**14.32 (9.77-18.48)**	**.41**	**< .001**		**14.95 (10.29-19.69)**	**.41**	**< .001**

*Note. N = 202, CI = confidence interval, significant models and predictors are highlighted in bold.*p<.01, **p<.001.*

Perceived support*:* As Table 6.5 above shows, all of the models predicting measures of perceived support were significant with proportion of times support is received when needed being the only significant predictor in all the models. This is the same pattern of results as seen in Study 1 and the variance accounted for by the models is comparable to that of the Step 2 models in Study 1. For emotional support, the average R^2 of the Step 2 models predicting perceived support outcomes in Study 1 was .34; in Study 2 it is .32. For tangible support, the average R^2 of the Step 2 models in Study 1 was .32; in Study 2 it is .26.

Health: Table 6.5 shows that, as in Study 1, the proportion of times support is received when needed is a significant predictor of mental but not physical health. In contrast to Study 1, age was a significant predictor of physical health but the regression models were not significant. In the two models predicting mental health scores all predictors apart from gender in the emotional support model were significant but the proportion of times support is received when needed made a greater contribution to both models than age and gender (emotional: $\beta_{age} = .26$, $\beta_{gender} = .18$, $\beta_{proportion} = .41$; tangible: $\beta_{age} = .25$, $\beta_{gender} = .20$, $\beta_{proportion} = .41$).

General Discussion

This research investigated whether the relationships between received and perceived support and received support and health were stronger when the need for support was taken into account. Study 1 showed that using a measure of received support that considers the need for support - the proportion of times support is received when needed - resulted in stronger correlations between received support and both perceived support and mental health measures than when received support was measured simply as the number of supportive behaviours received. Furthermore, the correlations between the proportion of times support is received when needed and both perceived support and mental health were even stronger when participants experiencing an oversupply of needed support were excluded from the analysis. This suggests that the positive relationship

between received support and both perceived support and mental health may break down in some way (i.e., become weaker, absent or even negative) when people receive more support than they need. Regression analyses supported these findings and a second study replicated the strength of the relationship between the proportion of times support is received when needed, perceived support and mental health.

Little evidence of a relationship between physical health and both received support and support received when needed was found but this may have been due to the way physical health was measured. Here we used the SF-36v2 physical component summary which is a general measure of functional health, i.e., the extent to which individuals perform regular, daily activities without limitations due to health problems. Previous studies investigating the relationship between social support and health have mainly looked at the association between social support and morbidity of or risk of mortality from specific chronic diseases (see Holt-Lunstad et al., 2010 for a review). Wolff et al. (2013) show a significant relationship between the balance of received and needed support and the experience of health complaints (e.g., headaches, upper respiratory complaints, muscle tension etc.). Our mental health measure was similar to this; unlike the physical component summary of the SF-36v2, the mental component summary comprised items asking about experiences of specific symptoms (e.g., feeling tired, low, nervous etc.) and we see much stronger effects with this measure than we do with physical health. It is likely that our results would have supported our hypotheses and the findings of Wolff et al. (2013) if we had measured morbidity of physical illness instead of functional health.

Furthermore, when we correlate scores from our two standardised measures of perceived support (MSPSS and ASSIS-S) with physical and mental health we find the same pattern of results from both studies, with the correlations between perceived support and physical health being weak and nonsignificant (Study 1: MSPSS: r = .12, ASSIS-S r = .13, Study 2: MSPSS: r = .03, ASSIS-S r = .08) but those between perceived support and mental

health being significant and much stronger (Study 1: MSPSS: *r* = .50, ASSIS-S *r* = .52, Study 2: MSPSS: *r* = .43, ASSIS-S *r* = .41). The strong relationship between perceived support and physical health has been well replicated further suggesting that the discrepancy in physical and mental health findings may have been due to the way physical health was measured.

Implications

Our findings suggest that measures such as the ISSB that are being used to investigate the relationship between received social support, perceived support and health need to ask not just about the amount of support received but about the amount of support received relative to the amount of times it is needed. Question wording such as that used in our second study where participants were asked to think about the last 50 occasions that they needed support and to state how many times they had actually received it may be more appropriate when investigating these relationships.

The findings may also have implications for stress buffering theory and interventions that have been developed based on this theory. Social support interventions are delivered in the same manner to all recipients. Our results suggest that increasing support may only have a beneficial effect on health when the recipient has identified a need for the support. Giving support when it is not needed or unwanted may have the opposite effect and may explain why interventions increasing support in an attempt to improve health have had mixed effects. Therefore, these interventions need to be individually tailored depending on support needs. Research investigating the efficacy of such interventions should control for the degree to which support needs were met prior to the intervention.

Limitations

Our study was limited in that we only looked at two different types of support (emotional and tangible) instead of all four identified by Barrera and Ainlay (1983). We could have taken items from the ISSB that measured directive guidance and positive social interactions as well. We chose to study only emotional and tangible support as we believed

that these types of support would be commonly received by all participants. Studies that have compared types of support typically only look at emotional and tangible support and some that have included directive guidance have found that this type of support was received infrequently by participants (e.g., Friedman & King, 1994).

Conclusions

The relationship between received and perceived support is affected by the need for support – people's perceptions relating to the support they receive are based not on the number of times they receive support but specifically on the number of times they have received it relative to the number of times they have needed it. The same is true for the relationship between received support and mental health. We have shown that perceptions of support become more positive and mental well-being increases as the percentage of times support needs are met increases to 100%. These relationships may break down beyond this, i.e., when people are given more support than they need.

Chapter 7: General Discussion

Overview

The studies presented in this thesis investigated the extent to which people compare to others when making a range of health-related judgements and decisions and aimed to identify the exact cognitive mechanisms used in this comparison process. All of the findings relating to these and other aims are discussed in the sections below along with the theoretical and applied implications of these results and potential areas for future research.

Main Findings

The Role of Social Comparison in Health-Related Judgement and Decision Making

One of the main aims of this thesis was to investigate whether social comparison effects could account for errors in health-related judgement and decision making such as inaccurate appraisals of symptoms, general health, sleep and social support and inaccurate health help-seeking decisions. In order for this to be ascertained, whether individuals actually compare to others when making such judgements and decisions needed to be established first. This was done by investigating whether measures of social comparison predicted the judgements and decisions of interest using regression analyses.

The results showed significant effects of how participants believed their experience of mental and physical health symptoms compared to other people's experiences of these symptoms. These beliefs were associated with judgements of: 1) symptom severity, 2) worry about symptoms, and 3) the presence of an underlying disorder, and decisions regarding: 1) whether to seek help for symptoms at all, 2) the number of help-seeking behaviours engaged with, 3) whether to consult a professional about symptoms and 4) whether to take medication for symptoms. Social comparison accounted for, on average,

15.4% (range: 6-28%) and 5.1%[10] (range: 1-11%) of the variance in these judgements and decisions respectively. Further analyses showed that the relationships between how participants believed their symptoms compared to others and decisions regarding whether to seek help for symptoms were mediated by judgements of symptom severity, worry about symptoms, and beliefs about an underlying disorder being present.

Participants' beliefs about how well they were generally in comparison to others accounted for around 6% of the variance in self-rated health (SRH) judgements. Even though the comparison effects seen in the study were small, when people were explicitly asked if they compared to others when making the judgement the vast majority of participants (84%) said that they did to some extent and, on average, participants who compared to others said that the comparison contributed 44% to the judgement.

Significant but much smaller social comparison effects were seen in the sleep study. Participants' beliefs about how their sleep compared to that of others accounted for only 1% of the variance in sleep-related worry and distress judgements and 2% and 4% of the variance in judgements of whether participants thought they had a sleep problem and sleep quality judgements respectively. There was little evidence of any associations between participants' beliefs about how their sleep compared to that of others and decisions regarding help-seeking.

No social comparison effects were seen at all in the first social support study; participants' beliefs about how the amount of support they received compared to the amount of support received by others did not significantly predict any of the five perceived support measures (measuring satisfaction with support, availability of support and satisfaction with the availability of support). Unfortunately, the independent variables were

[10] These and all of the other variance values described in this section are from the analyses using the rank comparison variables as they were much stronger predictors of the outcomes than the distance from the average comparison variables, as will be explained in the next section.

so highly correlated in the second study that social comparison effects could not be investigated.

In sum, the results provide evidence that social comparison plays a role both in general health and symptom appraisal and in help-seeking decision making for physical and mental health symptoms. Social comparison plays little role in the appraisal of sleep or making decisions about seeking help for sleep problems and there is no evidence that people compare to others when making judgements about the support they receive.

Cognitive Mechanisms Underlying Social Comparison

Another central aim of this thesis was to investigate how people compare to others when making health-related judgements and decisions as it is widely assumed (but untested) in the social comparison literature that people compare to the average of the comparison sample. Two measures of social comparison were used in all of the studies in order to ascertain whether people do compare to the sample average when making health-related judgements and decisions or whether they base these judgements and decisions on where they rank amongst the sample. The former is consistent with adaptation level theory (ALT) and the latter with decision by sampling (DbS).

Where the strongest comparison effects were seen, i.e., in the physical and mental health studies, there was overwhelming evidence that people compared to others using rank-based strategies and not by comparing to the sample average. The rank comparison variables (where participants believed their experience of symptoms ranked in comparison to other people's experiences of these symptoms) were significant predictors of all of the outcomes measured across all six symptoms - a total of 38 models. The average amount of variance accounted for by these variables was 9.1% (range: 1-28%). In contrast, the distance from the average variables (how much each participant's experience of the symptom differed from what they believe the average symptom experience of others to be) were only significant predictors in 12 out of the 38 models (32%) and accounted for, on

average, only 1.2% (range: 0-1%) of explained variance in these 12 models. In 37 out of the 38 full models, the step 2a models (including rank of symptom experience) either accounted for significantly more variance in the outcomes or there was very good evidence that the model fitted the data better than the step 2b models (including distance from the average symptom experience). Furthermore, in 31 out of the 38 step 2b models, rank was the strongest predictor of the outcomes, over and above participant's absolute experience of the symptom. When rank variables were added to the step 1 models that included age, gender, and the absolute experience, rank always either attenuated or removed the effect of the absolute experience.

The results from the general health study were inconclusive as to how participants compared to others when making SRH judgements. The correlations between the rank variables and SRH were significantly stronger than the correlations between the average variables. Furthermore, SRH and rank performed as well as objective health when predicting SRH judgements while the average was a weaker predictor. However, the model comparison showed that there was no difference in the fit of the step 2a and 2b models to the data. Although the social comparison effects in the sleep study were small, where there were significant differences in the fit of the step 2a and 2b models (9 out of 18 cases), it was the 2a rank model that fit the data better than the 2b distance from the average model.

In sum, the results generally suggest that people compare to others using rank-based strategies rather than comparing to the average of the sample when using social comparison to make health-related judgements and decisions.

Individual Differences in Beliefs about Others

In the studies where social comparison effects were consistently seen (mental health, physical health, and general health), further investigation into participants' beliefs about others was undertaken in order to examine whether social comparison effects may

explain inaccurate judgement and decision making. It was hypothesised that inaccurate judgements and decisions could be the result of people comparing to others when making the judgement/decision and using a comparison sample that was not representative of the actual state of the world. For example, a person who feels depressed 20 days a month could come to the decision that they do not need help if they make this decision by comparing their symptom occurrence to that of others who feel depressed more frequently than them. This is because, compared to these people, their symptom occurrence ranks low within the sample making it appear that they are better off than the majority of others. However, if that person had compared their symptom occurrence to the actual distribution of symptom occurrence in the general population then they may have concluded that they are in fact worse off than the majority of others and may have made the accurate decision to seek help. Similarly, inaccurate judgements/decisions may be made if an individual does compare to the general population but their beliefs about the distribution of what they are comparing are incorrect.

The results from all three health studies showed considerable variation in participants' beliefs about the general health of others and the occurrence and duration of various symptoms in the general population. This was examined through looking at the distributions of participants' answers to questions asking:

1. What they thought the average health of both people in their country and people their age was (general health study),
2. What they thought the average number of days in the last month (mental health study) or three months (physical health study) that people in the general population experienced feeling depressed and anxious (mental health study) and tired or run down and muscle, joint or back pain (physical health study) was and,
3. What they thought the average length of time that people in the general population experience headaches and colds for was (physical health study).

This variation indicates that many participants had incorrect beliefs about the general health of others and the frequency and duration with which people experience symptoms. Unsurprisingly, given these findings, there was also considerable variation in participants' beliefs about where their own symptom occurrence and duration ranked in comparison to others among participants who experienced the same symptom occurrence or duration. This indicates that many participants misestimated their rank. This could affect the accuracy of health-related judgements and decisions if participants base these judgements and decisions on their beliefs about how their symptoms compare to those of others. Interestingly, the correlation between the objective and self-rated health measures in the general health study was significantly stronger among participants who stated that they did not compare to others than among those that stated that they did. This suggests that participants who did not compare their health to others made more accurate judgements about their health than those who did.

When help-seeking accuracy could be examined in the mental health study, the findings showed that participants unlikely to have depression or anxiety were around four times more likely to seek help if they thought they experienced depression/anxiety symptoms more frequently than others. Participants likely to have clinical levels of depression or anxiety were two to three times more likely to not seek help if they thought they experienced symptoms less frequently than others.

Effect Sizes

The size of the comparison effects seen (based on the effects of the rank comparison variables) generally fell between the small and medium categories (average Cohen's f^2 = 0.09, range = 0.01-0.39). The largest effects were seen when judgements of headache and cold severity and worry about headache and cold duration were predicted by believed rank of headache and cold duration respectively (for headaches these were large effects and for colds they were medium effects).

As explained in Chapter 2, inclusion of participants' absolute experiences of the constructs measured (e.g., the number of days a month they felt depressed, the number of hours they slept, etc.) in the regression analyses provided a benchmark on which to compare the strength of the comparison effects. In the general health study, rank effects were similar to those of the absolute (the number of symptoms and conditions participants experienced). However, in the mental and physical health studies the rank variables were the strongest predictors of outcomes in the majority of the regression models (31 out of 38) and always either attenuated or removed the effect of the absolute variables (symptom occurrence/duration) when entered into the model. In the sleep study, the effects of the absolute variables (based on objective experiences of the five sleep aspects measured) were significantly larger than any comparison effects seen.

Implications of Main Findings and Areas for Future Research

These findings have a number of important implications, which are outlined below along with suggestions for future research directions.

Models of Symptom Appraisal and Help-seeking

The symptom appraisal process is complex and influenced by many person-related factors such as previous experience, beliefs and other cognitions, emotion, knowledge, personality, co-morbidity, and demographics (Scott, Walter, Webster, Sutton, & Emery, 2013; Whitaker et al., 2015). There is a consensus among symptom appraisal models that the appraisal process starts with the detection of bodily changes that are then interpreted and responded to (Whitaker et al., 2015). The interpretation process may involve labelling, categorising and evaluating the bodily change (e.g., symptom) such as judging whether it is severe or whether it maps to a known illness schema (Scott et al., 2013; Whitaker et al., 2015). The mental and physical health studies showed, for the first time, how social comparison might play an important role in the interpretation and appraisal of symptoms. Participants were found to appraise their symptoms using social comparison – their beliefs

about how their symptoms compared to others significantly predicted appraisals of symptom severity, worry about symptoms, and whether symptoms represented the presence of an underlying disorder. This provides a greater understanding of both the beliefs that influence symptom appraisal and the processes (i.e., rank-based comparison) involved in the evaluation of symptoms. This could be incorporated into existing models, for example, the common sense model of illness self-regulation (CSM; Leventhal, Meyer, & Nerenz, 1980). This model proposes that symptoms are appraised using heuristics such as the age-illness rule (the attribution of bodily changes to ageing rather than illness: Diefenbach & Leventhal, 1996) and the rate of change rule (sudden or worsening/ increasing/ unstable symptoms indicate illness and the need for care: Nyawata & Topping, 2006). Social comparison could be classed as such a heuristic in this model – judgement of a symptom experience to be worse/better than how the majority of others are believed to experience the symptom indicates the need to seek/not seek help. This proposition is supported by the work of Mussweiler and colleagues who have shown how quick and efficient comparative information processing can be. This has lead them to propose that social comparison is regularly used as a heuristic in judgement and decision-making (e.g., Mussweiler & Epstude, 2009; Mussweiler & Posten, 2012).

As previously mentioned, symptom appraisal models acknowledge that the interpretation process is followed by a response which might be the decision to take action, such as deciding to seek help from a professional, or to not act at all (Whitaker et al., 2015). Indeed, much research has shown that symptom appraisal plays a key role in help-seeking decision-making (Scott et al., 2013) and health beliefs/appraisals have long been recognised as influencers of behaviour in more general models of health behaviour (e.g., the theory of planned behaviour, Ajzen, 1988; the health belief model, Becker, 1974; protection motivation theory, Rogers, 1983). The current studies provide further evidence of the association between symptom appraisals and help seeking. Appraisals of symptom

severity, worry about symptoms, and whether symptoms represented the presence of an underlying disorder were all associated with whether or not participants sought help for symptoms. Furthermore, participants' beliefs about how their symptoms compared to others significantly predicted help seeking both directly and indirectly through these appraisals. The studies have therefore also identified another type of belief – how health is believed to compare to others – which influences help-seeking behaviour. However, perhaps the most important implication of these findings is that they can provide an understanding of how people might make inaccurate help seeking decisions.

In the context of help-seeking, inaccurate decisions are made when people either fail to seek help when they need to or seek help when they do not need to. As previously discussed, the majority of research in this area has focused on when and why the former occurs. The general model of total patient delay (Andersen et al., 1995; Safer et al., 1979) proposes that delay can occur at five time points between detecting abnormal sensations and beginning treatment for an illness. These are the periods between detecting sensations and evaluating whether they represent illness (appraisal delay), between making this judgement and deciding whether professional care is needed (illness delay), between making the decision to seek care and acting on it (behavioural delay), between acting and actually going to the clinic or service (scheduling delay), and between the first appointment with a professional and starting treatment (treatment delay). When these stages have been compared, appraisal delay has been found to contribute the most to the total delay time (e.g., Andersen et al., 1995; Ristvedt & Trinkaus, 2005; Walter, Webster, Scott, & Emery, 2011). This suggests that one of the main reasons why people are not seeking help when they need to is because they are making incorrect appraisals about their symptoms, for example, not judging them to be severe enough to warrant medical attention when in fact they are. Indeed, research has shown that errors in symptom appraisal such as

misattribution and failure to recognise the seriousness of symptoms contribute significantly to delays in help seeking (e.g., Evans, Ziebland, & McPherson, 2007).

Findings from the mental and physical health studies suggest that incorrect appraisals and help-seeking decisions may also be made when people's beliefs about how their symptoms compare to others are incorrect. For example, a person may judge that their symptoms are not severe enough to seek help if they believe that other people experience them more frequently than they do. Indeed, the studies showed an association between beliefs about how symptoms compare to others and help-seeking accuracy such that participants were more likely to have not sought help when they possibly needed to if they believed that they experienced symptoms less frequently than others. Similarly, participants were much more likely to seek help when they possibly did not need to if they believed that they experienced symptoms more frequently than others. Therefore, inaccurate health help-seeking decisions may be made if an individual bases their decision on how they believe their health compares to that of others but their beliefs about others are not representative of the actual state of the world.

The current studies may be the first to show the importance of social comparison in symptom appraisal and help-seeking. However, more research is needed to explore the role of social comparison further, using a wider variety of symptoms, aspects of symptom experience and other types of symptom appraisals, especially those known to affect help-seeking. The greater understanding of the beliefs and processes involved in symptom evaluation and appraisal that these studies have provided can also help to explain why people may make inaccurate judgements and decisions about their health and how such inaccuracies may be prevented (discussed in the next section). Inaccurate judgements and decisions may be made when people have inaccurate beliefs about the health of others. The studies showed that such inaccurate beliefs were prevalent but it is unknown as to whether certain people (i.e., those from a specific demographic group or with certain

personality types) may be more susceptible to holding inaccurate beliefs than others. For example, studies have shown that health-anxious individuals have different (i.e., more dysfunctional) assumptions and beliefs about health and illness compared to those with low levels of health anxiety (see Marcus, Gurley, Marchi, & Bauer, 2007 for a meta-analysis). Therefore, individuals high in health anxiety may be particularly likely to hold inaccurate beliefs about how their symptoms compare to others. Understanding whether there are any specific groups of people that may be more susceptible to holding inaccurate beliefs about others could help to identify those more likely to make inaccurate judgements and decisions about their health and therefore those who may be most effectively targeted with interventions to trying to improve accuracy in judgement and decision making. This could be a worthwhile area for future research.

Preventing Inaccuracy in Health-Related Judgement and Decision Making

As explained earlier, the results suggested that inaccurate judgements and decisions may be made if an individual bases their judgement/decision on how they believe their health compares to that of others but their beliefs about others are not representative of the actual state of the world. Such incorrect beliefs were prevalent in all three health studies and have been shown in other studies such as those investigating perceptions of health-related social norms and cognitive biases (e.g., the false consensus effect, pluralistic ignorance and unrealistic optimism) in these perceptions (see Suls, 2011, for a review). Given that such inaccurate beliefs about others may be quite prevalent, interventions that aim to recalibrate inaccurate beliefs may be effective in both encouraging people who need help to seek it and people that do not need help to refrain from seeking it. It has been suggested that providing people with the information needed to make accurate symptom appraisals via public health messages and education campaigns could improve help-seeking accuracy (Henshaw & Freedman-Doan, 2009). The results reported here suggest that such interventions could provide information about the actual distribution of aspects of

symptoms such as their occurrence and duration in order to encourage more accurate comparisons and hopefully more accurate decisions. Alternatively, objective information about symptom experiences (such as telling people to seek help if they have experienced a symptom for more than a certain number of days) could be provided in order to discourage social comparison and the errors in judgement that may occur because of it. Developing and testing the efficacy of such interventions could be a worthwhile area for future research.

Similar steps could be taken to help people make more accurate self-rated health judgements when they are answering these questions as part of population surveys or as clinical outcome measures. Previous research has shown that people often over- or underestimate their actual health when answering SRH questions. This finding has previously been explained as being the result of people comparing to others when making the SRH judgement and using different comparison groups when doing so. However, findings from the general health study suggest that people may still make inaccurate judgements about their health when using the same comparison sample (such as the general population) if their beliefs about the distribution of health in this sample are incorrect. It is not so much that people make inaccurate judgements about their health because they use different comparison groups; it is more about how representative these comparison groups are of the actual state of the world. It is therefore not enough to ask people to compare to a comparison group that is shared amongst all respondents when making the SRH judgement (which some surveys, such as BHPS, have done in an attempt to limit comparison effects). Surveys also need to ask respondents about their beliefs about the distribution of health in this comparison group so that their answers may be adjusted for varying beliefs. Alternatively, information about the actual distribution of health in the comparison group could be provided alongside the SRH question. This could be tested in

future research to see whether controlling for both comparison group and varying beliefs about others produces SRH judgements that more accurately reflect objective health.

Social Comparison Theory and Interventions

As explained in Chapter 1, although there has been a great deal of research on social comparison over the last sixty years it appears that the underlying mechanisms of the comparison process, i.e., how people actually compare to others, have received little attention in the literature to date. The results suggest that when people compare to others to make judgements and decisions about their health that they do so using rank-based strategies and not by comparing to the average of the sample as has previously been widely assumed. As hypothesised, this finding provides further support for the DbS model but not ALT and has a number of implications.

Firstly, this finding has implications for the development of social norm-based interventions and education campaigns which tend to supply people with information about how they compare to the average of a sample in the hope that people will modify their behaviour in response. The results suggest that these interventions may be more effective when rank-based information is supplied rather than information on how one differs from the average, particularly if the behaviour trying to be changed is health-related. Future research could investigate this further along with whether people compare to others using rank or average-based strategies when making self-assessments about other behaviours that are commonly targeted by social-norms based interventions such as recycling and energy consumption.

Secondly, this finding has implications for researchers investigating comparison effects by using large datasets to construct comparison groups. It should not be assumed that people compare to the average of others in the comparison group as has been done previously in such investigations (e.g., Powdthavee, 2009; Carrieri, 2012), as the individual's ranked position within the comparison group may be more influential. It should also be

noted that although data from large-scale studies can be used to construct comparison groups, the distribution of health in these groups might vary considerably from what people believe the distribution to be, which is what they would use to base the judgement about themselves on. Therefore, these studies may be missing comparison effects or underestimating their strength through using this methodology.

The physical and mental health studies showed social comparison effects that, to my knowledge, had not been explored before. Although it had previously been suggested that people might compare their experience of symptoms to that of others in order to make help-seeking decisions (Suls et al., 1997), it appears that this hypothesis has not been formally tested. Previous research on the use of social comparison in health-related decision-making has focused largely on the use of the lay referral structure for health advice with inaccuracy in help-seeking being explained as occurring due to receipt of inaccurate advice. The physical and mental health studies suggest that social comparison plays a much more pivotal role, not only affecting help-seeking decisions but also influencing symptom evaluation and appraisal.

Social comparison effects were also investigated, for what is believed to be the first time, in the domains of sleep and social support. Unfortunately, the comparison effects were so small in the sleep study that it was concluded that people rarely compared their sleep to that of others when appraising their sleep and making decisions regarding seeking help for sleep problems. Although there were some methodological limitations to the study (see Chapter 4), it was concluded that these probably had little effect on the estimations of the comparison effects. It is therefore possible that people simply do not compare their sleep that much to that of others when making sleep-related judgements and decisions. The general health study reported in Chapter 5 showed that people often compared to their previous health when making judgements about their current state of health. Perhaps

then, if people base judgements and decisions about their sleep on a comparison, it is to their previous sleep experiences rather than other people's sleep experiences.

No comparison effects were found in the first social support study suggesting that participants' perceptions of the support they receive are not associated with how they believe the amount of support they receives compares to others. Unfortunately, this could not be investigated in the second study, which utilised a better methodology that controlled for the need for support, as the independent variables were so highly correlated. It would have been interesting to see if any comparison effects would have been present if the new, direct rank and distance from the average comparison measures had been used in those studies. Overall, the findings from the sleep and social support studies suggest that these are not fruitful areas for social comparison research or interventions.

Other Findings

There were a number of other findings from the general health and social support studies that were not directly related to the central aims of the thesis and these are summarised briefly next.

Individual Differences in the use of Health Aspects and Comparison

The general health study also investigated what aspects of their health participants used when making the SRH judgement and when making comparisons to others and their previous health, the extent to which participants based their SRH judgement on both how their health compares to that of others and to their own previous health, and which comparison groups participants used when making SRH judgements.

General physical condition was the health aspect that was used the most when making SRH judgements (27.8%) and when comparing current health to both others (12.5%) and previous health (11.5%). The second and third most important aspects considered when making the SRH judgement were the presence (23.3%) and absence (12.6%) of health problems. The second and third most important aspects considered when

comparing to others were engagement with healthy behaviours (12%) and energy (11.2%) and when comparing to previous health were general feeling (11.1%) and energy (9.8%). The data suggested that the health aspects that participants most commonly compared to others and to their previous health when judging their general health were very similar. In the majority of cases, participants were no more likely to use one comparison type (others vs previous health) over the other when using a specific health aspect to make the SRH judgement.

When asked explicitly, the majority of participants said that they compared their current health to that of others (84%) and/or their previous health (90%) to some extent when answering the SRH question. On average, participants said that the comparison to others contributed 44% to the judgement and the comparison to previous health contributed 47%. Women and participants over 40 based a higher percentage of their SRH judgement on how their current health compared to their previous health than men and participants under 30 respectively. These findings provide evidence of further context effects that may affect SRH judgements.

When making the SRH judgement, participants tended to compare to people that were around them: friends (23%), family (19.3%) and people around their age (18.4%), although quite a few referred to the health of the general population (13.8%). Differences in the choice of comparison group have largely been unexplored in the previous literature and were therefore investigated here. Both younger and healthy participants were more likely than older participants and those in poor health to compare to family members. Older participants were more likely than younger to compare to people their age and participants in poor health were more likely than healthy to compare to people their age and the general population. These findings confirm that people use different comparison groups when evaluating their health and that there are systematic differences in the use of comparison group.

Measurement of Received Support

As reported in Chapter 6, when undertaking the social support studies a novel finding was uncovered: the relationship between the amount of support a person receives (received support) and their perceptions of this support (perceived support) is affected by whether the person actually needed the support they received. When the need for support was taken into consideration, correlations between received and perceived support increased substantially. This finding explained why such low correlations between these two constructs had been found previously, despite theoretical accounts suggesting that there should be a strong relationship between the two. One measure of received support had been consistently used in the investigation of the received-perceived support relationship and this did not control for the need for support. Therefore, the main implication of this finding was that measures of received support need to ask about both the amount of support received and whether it was needed.

General Limitations

In addition to the limitations exclusive to each study (outlined in the general discussions of each chapter), there were a number of general limitations to the research presented in this thesis. Ideally, objective health, the presence of clinical levels of depression, anxiety and insomnia and help seeking behaviours undertaken for these disorders (particularly whether participants had consulted a professional or taken medication) would have been ascertained through methods that may have been more reliable and valid than self-report, such as clinical interviews or medical record screening. However, this was not possible due to the large number of participants required for each study. Recruiting and medically screening thousands of participants would not have been feasible due to time and monetary constraints. Whilst collecting information on symptom presence and help-seeking from participants is much easier and cheaper, this information can be inaccurate due to factors such as willingness to report, recall timeframe, frequency

of help seeking, questionnaire design, participants' knowledge/understanding of what they are being asked to report and recall ability (Goldman et al., 2003; Bhandari & Wagner, 2006). However, the design of the questionnaires may have helped to reduce the likelihood of these factors affecting the reliability and validity of the self-reported measures. For example, the questionnaires were all online surveys (i.e., no face-to-face or telephone contact with the experimenter), no identifiable information was requested and participants were told that their responses were confidential. This may have made participants more willing to report symptoms that they may be embarrassed about or feel are stigmatised and may have encouraged them not to under-report help seeking. The longer the recall timeframe provided, the worse recall accuracy becomes. Bhandari and Wagner (2006) recommend avoiding recall timeframes greater than 12 months which was adhered to in the current studies (timeframes were mainly three and six months). Recall accuracy is also affected by the number of times people engage in the help seeking behaviour (Bhandari & Wagner, 2006). However, this is largely concerned with recalling the frequency of engagement – for example, the more frequently people visit their doctor the more likely they are to under-report the number of doctor visits simply because they forget some. The questionnaires asked only whether participants engaged in each type of help-seeking behaviour and not how many times they did so and so this is not likely to be an issue in the current studies. The questionnaires were worded as clearly as possible to avoid any misunderstanding of what was being asked. For example, participants were not just asked whether they had consulted a professional, they were specifically asked whether they had consulted a doctor, nurse, pharmacist, or therapist. As explained in Chapter 5, research comparing self-report and medical report of symptoms has shown that people tend to be good at accurately reporting whether or not they have many of the symptoms/illnesses presented in the measure of objective health in the general health study. The depression and anxiety screening measures (PHQ-9 and GAD-7) were standardised measures validated

for use within the general population with sensitivity and specificity of over 80%. Although the self-report insomnia screening measure was developed for the sleep study (as no standardised measure reflecting the current diagnostic criteria for insomnia was available at the time), this was not used in any analyses.

A second limitation of the research is that no conclusions can be made about the causal relationships between variables, as all of the studies were correlational in design. However, these were exploratory studies - one of the main aims of the research was to see whether the relationships studied actually existed, which was not the case in some circumstances (i.e., the sleep and social support studies). Where relationships between social comparison variables and outcomes have been shown to exist, a natural next step for this research is to examine causality using an experimental design. Rank has previously been manipulated experimentally using a between subjects design and by presenting participants with one of two distributions of values, for example, the number of days a month 11 people in the general population feel depressed (Melrose et al., 2013). The two distributions have the same mean and end points and three common values in addition to the end points, for example: distribution 1: 3, 10, 12, 13, 14, 16, 18, 19, 20, 22, and 29 days, distribution 2: 3, 4, 6, 8, 10, 16, 22, 24, 26, 28, and 29 days. The common value of 16 (the mean) is ranked 6th in both distributions but the other two common values (10 and 22) appear at different ranked positions within the two distributions – 10 is ranked lower in distribution 1 than in distribution 2 (2nd compared to 5th) and 22 is ranked higher in distribution 1 than in distribution 2 (10th compared to 7th). Participants are asked to make a judgement about the values, such as how severe they think the symptom occurrence is on a given scale. Studies utilising this paradigm such as Melrose et al. (2013) tend to show that there is no significant difference in the average rating given by participants in the two groups to the value which is ranked the same in both distributions (i.e., the value of 16 in the above example). However, the average ratings given to the other two common values

do differ significantly between the groups. In the above example, participants given distribution 2 rated the occurrence of 10 days a month to be more severe than participants given distribution 1. The opposite was true for the occurrence of 22 days. This suggests that the ranked position of the number of days affected how severe it was perceived to be by the participants.

Thirdly, as with any questionnaire based study there is the possibility that item order effects may have occurred. However, care was taken to overcome any potential effects through counterbalancing the presentation order of items. In all of the studies apart from the general health study, questions were split into two blocks of independent variable questions (absolute experience, rank, and average questions) and dependent variable questions (all other questions such as help seeking, severity, and worry questions). Half of the participants in each study saw the independent variable question block first and the other half saw the dependent variable question block first. In addition, where the order of the independent variable questions was not imperative to being able to answer the questions (i.e., in all but the mental health study where the average questions had to follow the rank questions because of the way they were worded), the presentation order of the absolute, rank and average questions was randomised within the question block. For the sleep study, the presentation order of questions within the dependent variable block was also randomised (later studies had more sophisticated counterbalancing as the online survey software developed and improved over time). For the general health study, all participants answered the self-rated health question first and then the presentation order of the rank and average questions was counterbalanced so that half of the participants saw the rank questions first and half saw the average questions first. The presentation order of the remaining questions was the same for all participants. This was because all the other questions related to how participants answered the self-rated health question and therefore they had to follow on from this question.

Finally, as explained in Chapter 1, people compare to others when faced with uncertainty (Festinger, 1954; Schachter & Singer, 1962). Therefore, the use of social comparison in health-related judgement and decision-making may be dependent on how uncertain the person is about the judgement/decision they are making. For example, a person who has suffered from depression for many years will know to seek help if they start to experience an episode. It is unlikely that they will compare their symptoms to those of others in order to make this decision because they are likely to recognise how they are feeling from past experience and be quite certain in how they should address their symptoms. However, a person experiencing depression for the first time may be quite uncertain about how they feel and whether they should seek help and therefore more likely to use social comparison to help make this decision. It is therefore likely that the extent to which each participant used social comparison when making the judgements and decisions studied differed according to their uncertainty. However, this was not captured in the analyses, which show an average comparison effect across all participants. In future studies it would be interesting to measure judgement/decision making uncertainty and investigate whether social comparison effects are stronger in participants who are more uncertain and weaker in those who are more certain as is hypothesised.

Conclusions

The studies presented in this thesis showed that participants compared to others when making judgements and decisions regarding their general health and both somatic and psychological symptoms but not their sleep nor the support they receive from others. When participants compared to others in order to make health-related judgements and decisions they did so using their rank within the comparison sample and not their distance from the sample mean, which has often been assumed in previous social comparison research. Participants' beliefs about the distributions of general health, symptom occurrence, and symptom duration in the general population varied considerably,

indicating that many participants had incorrect beliefs about the health and symptom experience of others.

When such inaccurate beliefs about others are used to make judgements and decisions about one's own health through social comparison, it is likely that these judgements and decisions will also be inaccurate. The physical and mental health studies provided evidence of this. Estimates of where their symptom occurrence and duration ranked in comparison to others given by participants who experienced the same symptom occurrence or duration varied substantially, indicating that many participants misestimated their rank. Given that the studies showed an association between believed rank of symptom experience and help-seeking, it is likely that such misestimations may have led to inaccurate help-seeking decisions. Indeed, help-seeking inaccuracy was associated with beliefs about how symptoms compared to others. Participants unlikely to have depression or anxiety were around four times more likely to seek help if they thought they experienced depression/anxiety symptoms more frequently than others (i.e., ranked high within the comparison sample). Participants likely to have clinical levels of depression or anxiety were two to three times more likely not to seek help if they thought they experienced symptoms less frequently than others (i.e., ranked low within the comparison sample).

These findings provide a better understanding of how people evaluate their general health and symptoms and make judgements and decisions about them. They suggest that inaccurate health help-seeking behaviours (and possibly other inaccurate health-related judgements and decisions) may be targeted effectively by interventions that aim to recalibrate incorrect beliefs about other people's health experiences, for example, by providing information about the actual distribution of symptom experience, so that people can make more accurate self-evaluations. Future research could test whether such interventions are effective in encouraging people to seek help for symptoms or disorders

that research has shown people often do not seek help for when they need to (such as mental health disorders) and seek help when they do not need to (such as cold and flu symptoms).

This research investigated social comparison effects in new contexts and the suggestion that people may compare to others using rank-based strategies has implications for social comparison theory, research and interventions. More research is needed to ascertain whether people use rank-based strategies globally in social comparison or whether this is specific to health comparisons. Research is also needed to investigate whether people use rank-based strategies when making self-assessments about behaviours commonly targeted by social-norms based interventions. The effects of such interventions may have been underestimated previously because these interventions have given feedback based on how the individual compares to the average of the comparison sample and not where they rank within the sample.

Chapter 8: References

Adler, D. A., McLaughlin, T. J., Rogers, W. H., Chang, H., Lapitsky, L., & Lerner, D. (2006). Job performance deficits due to depression. *American Journal of Psychiatry, 163,* 1569–1576.

Agostinelli, G., Brown, J., & Miller, W. R. (1995). Effects of normative feedback on consumption among heavy drinking college students. *Journal of Drug Education, 25*(1), 31–40.

Agresti, A. (1996). *An Introduction to Categorical Data Analysis.* New York, NY: Wiley.

Ajzen, I. (1988). *Attitudes, personality & behavior*. Milton Keynes, UK: Open University Press.

Aldrovandi, S., Wood, A. M., & Brown, G. D. A. (2013). Sentencing, severity, and social norms: A rank-based model of contextual influence on judgments of crimes and punishments. *Acta Psychologica, 144,* 538–547.

Aldrovandi, S., Wood, A. M., Maltby, J. & Brown, G. D. A. (2015). Students' concern about indebtedness: a rank based social norms account. *Studies in Higher Education*, *40*(7), 1307-1327.

Alicke, M. (1985). Global self-evaluations as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49,* 1621–1630.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed). Washington, DC: Author.

Ancoli-Israel, S., & Roth, T. (1999). Characteristics of insomnia in the United States: results of the 1991 national sleep foundation survey. *Sleep, 22,* 347–353.

Andersen, B. L., Cacioppo, J. T., & Roberts, D. C. (1995). Delay in seeking a cancer diagnosis: Delay stages and psychophysiological comparison processes. *British Journal of Social Psychology, 34,* 33–52.

Andrews, G., Issakidis, C., & Carter, G. (2001). Shortfall in mental health service utilisation. *The British Journal of Psychiatry, 179,* 417–425.

Arigo, D., Suls, J. M., & Smyth, J. M. (2014). Social comparisons and chronic illness: research synthesis and clinical implications. *Health Psychology Review, 8*(2), 154-214.

Arnault, D. S. (2009). Cultural determinants of help-seeking: A model for research and practice. *Research and Theory for Nursing Practice: An International Journal, 23,* 259–277.

Au, N., & Johnston, D. W. (2013). An econometric analysis of self-assessed health: what does it mean and what is it hiding? Working Paper No. 13/31. HEDG, Department of Economics, University of York.

Backett, E. H., Heady, J. A., & Evans, J. C. G. (1954). Studies of general practice. The doctor's job in an urban area. *British Medical Journal, 1,* 109-115.

Barbara, A. M., Loeb, M., Dolovich, L., Brazil, K., & Russell, M. (2012). Agreement between self-report and medical records on signs and symptoms of respiratory illness. *Primary Care Respiratory Journal, 21*(2), 145-152.

Barber, J., Muller, S., Whitehurst, T., & Hay, E. (2010). Measuring morbidity: self-report or health care records? *Family Practice, 27(*1), 25-30.

Baron-Epel, O., Kaplan, G., Haviv-Messika, A., Tarabeia, J., Green, M. S., & Kaluski, D. N. (2005). Self-reported health as a cultural health determinant in Arab and Jewish Israelis MABAT – National Health and Nutrition Survey 1999–2001. *Social Science & Medicine, 61*(6), 1256–1266.

Barrera, M., Jr. (1980). A method for the assessment of social support networks in community survey research. *Connections, 3*(3), 8-14.

Barrera, M., Jr. (1981). Social support in the adjustment of pregnant adolescents. In B. H. Gottlieb (Ed.), *Social networks and social support* (pp. 69-96). Beverly Hills, CA: Sage.

Barrera, M., Jr. (1986). Distinctions between social support concepts, measures and models. *American Journal of Community Psychology, 14,* 413–455.

Barrera, M., Jr. (2000). Social support research in community psychology. In J. Rappaport & E. Seidman (Eds.), *Handbook of community psychology* (pp. 215–246). New York, NY: Plenum.

Barrera, M., Jr., & Ainlay, S. L. (1983). The structure of social support: A conceptual and empirical analysis. *Journal of Community Psychology, 11,* 133-143.

Barrera, M., Jr., Glasgow, R. E., McKay, H. G., Boles, S. M., & Feil, E. G. (2002). Do internet-based support interventions change perceptions of social support? An experimental trial of approaches for supporting diabetes self-management. *American Journal of Community Psychology, 30,* 637–654.

Barrera, M., Jr., Sandler, I. N., & Ramsay, T. B. (1981). Preliminary development of a scale of social support: studies on college students. *Journal of Community Psychology, 9,* 435-447.

Barsky, A. J. (1981). Hidden reasons some patients visit doctors. *Annals of Internal Medicine, 94,* 492-498.

Bastien, C. H., Vallières, A., Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine, 2,* 297-307.

Baumeister, H., Kriston, L., Bengel, J., & Härter, M. (2010). High agreement of self-report and physician-diagnosed somatic conditions yields limited bias in examining mental-physical comorbidity. *Journal of Clinical Epidemiology, 63,* 558-565.

Becker, M. H. (1974). The health belief model and sick role behaviour. *Health Education Monographs,* 2, 409–419.

Benyamini, Y., & Idler, E. L. (1999). Community studies reporting association between self-rated health and mortality: Additional studies, 1995 to 1998. *Research on Aging, 21,* 392-401.

Benyamini, Y., Idler, E. L., Leventhal, H., & Leventhal, E. A. (2000). Positive affect and function as influences on self-assessments of health: Expanding our view beyond illness and disability. *Journals of Gerontology. Series B: Psychological Sciences and Social Sciences, 55*(2), 107–116.

Benyamini, Y., Leventhal, E. A., & Leventhal, H. (1999). Self assessments of health: What do people know that predicts their mortality? *Research on Aging, 21*(3), 477–500.

Benyamini, Y., Leventhal, E. A., & Leventhal, H. (2003). Elderly people's ratings of the importance of health-related factors to their self-assessments of health. *Social Science and Medicine, 56,* 1661–1667.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon. com's Mechanical Turk. *Political Analysis, 20,* 351–368.

Berkanovic, E., Telesky, C., & Reader, S. (1981). Structural and social psychological factors in the decision to seek medical care for symptoms. *Medical Care, 19,* 693-709.

Bhandari, A., & Wagner, T. (2006). Self-reported utilization of health care services: Improving measurement and accuracy. *Medical Care Research and Review, 63*(2), 217-235.

Biernat, M., & Billings, L. S. (2001). Standards, expectancies and social comparison. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 257-283). Malden, MA: Blackwell Publishers Inc.

Birnbaum, M. (1974). Using contextual effects to derive psychophysical scales. *Perception and Psychophysics, 15,* 89-96.

Bolger, N., & Amarel, D. (2007). Effects of social support visibility on adjustment to stress: Experimental evidence. *Journal of Personality and Social Psychology, 92*(3), 458-475.

Bonnet, M., & Arand, D. (1997). Hyperarousal and insomnia. *Sleep Medicine Reviews, 2,* 97-108.

Bootzin, R. R. (1972). Stimulus control treatment for insomnia. *Proceedings of the American Psychological Association, 7,* 395–396.

Braun, B. L., Fowles, J. B., Solberg, L., Kind, E., Healey, M., & Anderson, R. (2000). Patient beliefs about the characteristics, causes, and care of the common cold. *Journal of Family Practice, 49,* 153–156.

Breslau, N., Roth, T., Rosenthal, L., & Andreski, P. (1996). Sleep disturbance and psychiatric disorders: A longitudinal epidemiological study of young adults. *Biological Psychiatry, 39,* 411–418.

Brown, G. D. A., & Matthews, W. J. (2011). Decision by sampling and memory distinctiveness: range effects from rank-based models of judgment and choice. *Frontiers in Psychology, 2,* 299.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114,* 539–576.

Brown, G. D. A., Wood, A. M., Ogden, R. S., & Maltby, J. (2015). Do student evaluations of university reflect inaccurate beliefs or actual experience? A relative rank model. *Journal of Behavioral Decision Making, 28,* 14–26.

Bruck, D., Dolan, C. L., & Lack, L. C. (2015). Beliefs about the 'shape' and continuity of healthy sleep as a function of age. *Journal of Psychosomatic Research, 78,* 39-44.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research, 33*(2), 261-304.

Buunk, B. P., Gibbons, F. X., & Visser, A. (2002). The relevance of social comparison processes for prevention and health care. *Patient Education and Counseling, 47,* 1–3.

Buysse, D. J., Angst, J., Gamma, A., Ajdacic, V., Eich, D., & Rössler, W. (2008). Prevalence, course, and comorbidity of insomnia and depression in young adults. *Sleep, 31*(4), 473–480.

Buysse, D. J., Germain, A., Hall, M., Monk, T. H., & Nofzinger, E. A. (2011). A neurobiological model of insomnia. *Drug Discovery Today. Disease Models, 8*(4), 129-137.

Buysse, D., Reynolds, C. F., Monk, T., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatric Research, 28*(2), 193-213.

Cameron, L., Leventhal, E. A., & Leventhal, H. (1993). Symptom representation and affect as determinants of care seeking in community-dwelling, adult sample, population, *Health Psychology, 12,* 171-179.

Cappuccio, F. P., Cooper, D., D'Elia, L., Strazzullo, P., & Miller, M. A. (2011). Sleep duration predicts cardiovascular outcomes: A systematic review and meta-analysis of prospective studies. *European Heart Journal, 32,* 1484–1492.

Carney, C. E., & Edinger, J. D. (2006). Identifying critical dysfunctional beliefs about sleep in primary insomnia. *Sleep, 29,* 440–453.

Carney, C. E., Edinger, J. D., Manber, R., Garson, C., & Segal, Z. V. (2007). Beliefs about sleep in disorders characterized by sleep and mood disturbance. *Journal of Psychosomatic Research, 62,* 179–188.

Carrieri, V. (2012). Social comparison and subjective well-being: Does the health of others matter? *Bulletin of Economic Research, 64*(1), 31-55.

Cauce, A., Domenech-Rodriguez, M., Paradise, M., Cochran, B. N., Shea, J. M., Srebnik, D., & Baydar, N. (2002). Cultural and contextual influences in mental health help-seeking: A focus on ethnic minority youth. *Journal of Consulting and Clinical Psychology, 70*(1), 44–55.

Cheung, J. M. Y., Bartlett, D. J., Armour, C. L., Glozier, N., & Saini, B. (2014). Insomnia patients' help-seeking experiences. *Behavioral Sleep Medicine, 12*(2), 106-122.

Chong, Y., Fryar, C. D., & Gu, Q. (2013). Prescription sleep aid use among adults: United States, 2005-2010. *NCHS Data Brief, 127,* 1-8.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin, 98,* 310–57.

Collins, R. L. (1996). For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin, 119,* 51-69.

Colten, H. R., Altevogt, B. M., & the Institute of Medicine Committee on Sleep Medicine and Research. (2006). *Sleep disorders and sleep deprivation: An unmet public health problem.* Washington, DC: Institute of Medicine/National Academies Press.

Contoyannis, P., Jones, A. M., & Rice, N. (2004). The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics, 19,* 473-503.

Corcoran, K., & Mussweiler, T. (2010). The cognitive miser's perspective: Social comparison as a heuristic in self-judgements. *European Review of Social Psychology, 21*(1), 78-113.

Cornally, N., & McCarthy, G. (2011). Help-seeking behaviour: A concept analysis. *International Journal of Nursing Practice, 17,* 280–288.

Costa, P. T., & McCrae, R. R. (1980). Somatic complaints in males as a function of age and neuroticism: A longitudinal analysis. *Journal of Behavioural Medicine, 3,* 245-257.

Cott, C. A., Gignac, M. A. M., & Badley, E. M. (1999). Determinants of self-rated health for Canadians with chronic disease and disability. *Journal of Epidemiology and Community Health, 53*(11), 731–736.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96*(4), 608–630.

Cumming, G. (2009). Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine, 28,* 205–220.

Cutrona, C. E., & Russell, D. W. (1990). Type of social support and specific stress: Toward a theory of optimal matching. In B. R. Sarason, I. G. Sarason, & G. R. Pierce (Eds.), *Social support: An interactional view* (pp. 319–366). New York, NY: Wiley.

Daley, M., Morin, C. M., LeBlanc, M., Grégoire, J. P., Savard, J., & Baillargeon, L. (2009) Insomnia and its relationship to health-care utilization, work absenteeism, productivity and accidents. *Sleep Medicine, 10,* 427–438.

Davis, B., Moore, B., & Bruck, D. (2006). The meanings of sleep: Stories from older women in care. *Sociological Research Online, 12*(5), 7.

DeSalvo, K. B., & Muntner, P. (2011). Discordance between physician and patient self-rated health and all-cause mortality. *The Ochsner Journal, 11,* 232–240.

Diefenbach, M. A., & Leventhal, H. (1996). The common-sense model of illness representation: Theoretical and practical considerations. *Journal of Social Distress and the Homeless, 5*(1), 11–38.

Dolan, P., & Metcalf, R. (2013). Neighbors, knowledge, and nuggets: Two natural field experiments on the role of incentives on energy conservation. *Centre for Economic Performance Discussion Paper No. 1222.*

van Doorslaer, E., Jones, A. M., & Koolman, X. (2004). Explaining income-related inequalities in doctor utilization in Europe. *Health Economics, 13,* 629–647.

Druss, B. G., Rosenheck, R. A., & Sledge, W. H. (2000). Health and disability costs of depressive illness in a major U.S. corporation. *American Journal of Psychiatry, 157,* 1274–1278.

Edinger, J. D., Wohlgemuth, W. K., Radtke, R. A., Marsh, G. R., & Quillian, R. E. (2001a). Cognitive behavioral therapy for treatment of chronic primary insomnia: a randomized controlled trial. *Journal of the American Medical Association, 285*(14), 1856–1864.

Edinger, J. D., Wohlgemuth, W. K., Radtke, R. A., Marsh, G. R., & Quillian, R. E. (2001b). Does cognitive-behavioral insomnia therapy alter dysfunctional beliefs about sleep? *Sleep, 24*(5), 591–599.

Elliott, A. M., McAteer, A., & Hannaford, P. C. (2011). Revisiting the symptom iceberg in today's primary care: results from a UK population survey. *BMC Family Practice, 12,* 16.

Ellis, J., Gehrman, P., Espie, C., Riemann, D., & Perlis, M. (2012). Acute insomnia: Current conceptualizations and future directions. *Sleep Medicine Reviews, 16*(1), 5-14.

Ellis, J., Perlis, M., Neale, L., Espie, C., & Bastien, C. (2012). The natural history of insomnia: Focus on prevalence and incidence of acute insomnia. *Journal of Psychiatric Research, 46*(10), 1278-1285.

Espie, C. A., Broomfield, N. M., MacMahon, K. M., Macphee, L. M., & Taylor, L. M. (2006). The attention-intention-effort pathway in the development of psychophysiologic insomnia: a theoretical review. *Sleep Medicine Reviews, 10*(4), 215-245.

Espie, C. A., Inglis, S. J., Harvey, L., & Tessier, S. (2000). Insomniacs' attributions: psychometric properties of the dysfunctional beliefs and attitudes about sleep scale and the sleep disturbance questionnaire. *Journal of Psychosomatic Research, 48,* 141–148.

Ettner, S. L. (1996). New evidence on the relationship between income and health. *Journal of Health Economics, 15,* 67-85.

Evans, J., Ziebland, S., & McPherson, A. (2007). Minimizing delays in ovarian cancer diagnosis: An expansion of Andersen's model of "total patient delay. *Family Practice, 24*(1), 48–55.

Facione, N. C. (1993). Delay versus help-seeking for breast cancer symptoms: A critical review of the literature on patient and provider delay. *Social Science and Medicine, 36,* 1521–1534.

Fayers, P. M., & Sprangers, M. A. G. (2002). Understanding self-rated health. *The Lancet, 359*(9302), 187-188.

Fayers, P. M., Langston, A. L., Robertson, C., & PRISM Trial Group. (2007). Implicit self-comparisons against others could bias quality of life assessments. *Journal of Clinical Epidemiology, 60*(10), 1034–1039.

Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-attribute health status classification systems: Health Utilities Index. *Pharmacoeconomics, 7*(6), 490-502.

Ferraro, K. F., & Farmer, M. M. (1999). Utility of health data from social surveys: Is there a gold standard for measuring morbidity? *American Sociological Review, 64*(2), 303-315.

Ferraro, K. F., Farmer, M. M., & Wybraniec, J. A. (1997). Health trajectories: Long-term dynamics among Black and White adults. *Journal of Health and Social Behavior, 38*(1), 38-54.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117-140.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review, 107*(4), 659-676.

Ford, D. E, & Kamerow, D. B. (1989). Epidemiologic study of sleep disturbances and psychiatric disorders. An opportunity for prevention? *JAMA, 262,* 1479–1484.

Ford, J., Spallek, M., & Dobson, A. (2008). Self-rated health and a healthy lifestyle are the most important predictors of survival in elderly women. *Age and Ageing, 37*(2), 194–200.

Freidson, E. (1961). *Patients' view of medical practice: A study of subscribers to a prepaid medical plan in the Bronx.* New York, NY: Russell Sage Foundation.

Friedman, M. M., & King, K. B. (1994). The relationship of emotional and tangible support to psychological well-being among older women with heart failure. *Research in Nursing and Health, 17,* 433-440.

Friedman, O., Bradley, T. D., Ruttanaumpawan, P., & Logan, A. G. (2010). Independent association of drug-resistant hypertension to reduced sleep duration and efficiency. *American Journal of Hypertension, 23,* 174–179.

Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine, 33,* 357–362.

Geest, T. A., Engberg, M., & Lauritzen, T. (2004). Discordance between self-evaluated health and doctor-evaluated health in relation to general health promotion. *Scandinavian Journal of Primary Health Care, 22*(3), 146-151.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65,* 263–297.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225–277). London: Cambridge University Press.

Gilbert, D. T., Giesler, R. B., & Morris, K. A. (1995). When comparisons arise. *Journal of Personality and Social Psychology, 69,* 227-236.

Goethals, G. R., & Darley, J. M. (1977). Social comparison theory: An attributional approach. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 259-278). Washington, DC: Hemisphere.

Goff, D. C., Nichaman, M. Z., Ramsey, D. J., Meyer, P. S., & Labarthe, D. R. (1995). A population-based assessment of the use and effectiveness of thrombolytic therapy: The Corpus Christi Heart Project. *Annals of Epidemiology, 5,* 5171-5178.

Goff, D. C., Sellers, D. E., McGovern, P. G., Meischke, H., Goldberg, R. J., Bittner, V., … Nichaman, M. Z. (1998). Knowledge of heart attack symptoms in a population survey in the United States. *Archives of Internal Medicine, 158,* 2329-2338.

Goldman, N., Lin, I. F., Weinstein, M., & Lin, Y. H. (2003). Evaluating the quality of self-reports of hypertension and diabetes. *Journal of Clinical Epidemiology, 56*(2), 148-54.

Goldsmith, H. F., Jackson, D. F., & Hough, R. L. (1998). Process model of seeking mental health services: Proposed framework for organizing the research literature on help-seeking. In: H. F. Goldsmith (Ed.), *Needs assessment: It's future* (pp. 49-93). University of Michigan, MI: National Institute of Mental Health.

Gooneratne, N. S., Richards, K. C., Joffe, M., Lam, R. W., Pack, F., Staley, B., … Pack, A. I. (2011). Sleep disordered breathing with excessive daytime sleepiness is a risk factor for mortality in older adults. *Sleep, 34,* 435–442.

Gottlieb, B. H., & Bergen, A. E. (2010). Social support concepts and measures. *Journal of Psychosomatic Research, 69,* 511-520.

Grandner, M. A., Patel, N. P., Hale, L., & Moore, M. (2010). Mortality associated with sleep duration: The evidence, the possible mechanisms, and the future. *Sleep Medicine Reviews, 14,* 191–203.

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics, 19,* 403-420.

Haber, M. G., Cohen, J. L., Lucas, T., & Baltes, B. B. (2007). The relationship between self-reported received and perceived social support: A meta-analytic review. *American Journal of Community Psychology, 39,* 133–144.

Harries, T., Rettie, R., Studley, M., Burchell, K., & Chambers, S. (2013). Is social norms marketing effective? A case study in domestic electricity consumption. *European Journal of Marketing, 47*(9), 1458-1475.

Harvey, A. G. (2002). A cognitive model of insomnia. *Behaviour Research and Therapy, 40,* 869–893.

Harvey, A. G., Stinson, K., Whitaker, K. L., Moskovitz, D., & Virk, H. (2008). The subjective meaning of sleep quality: a comparison of individuals with and without insomnia. *Sleep, 31*(3), 383-393.

Hauser, D. J., & Schwarz, N. (2015). Attentive Turkers: MTurk participants perform better on online attention checks than subject pool participants. *Behavior Research Methods.* Advance online publication.

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. Retrieved from http://www.afhayes.com/public/process2012.pdf

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st Century. *Medical Care, 38*(9 Suppl.), II28- II42.

Hedges, J. R., Mann, N. C., Meischke, H., Robbins, M., Goldberg, R., & Zapka, J. (1998). Assessment of chest pain onset and out-of-hospital delay using standardized interview questions: the REACT Pilot Study. Rapid Early Action for Coronary Treatment (REACT) Study Group. *Academic Emergency Medicine, 5*(8), 773-780.

Helson, H. (1947). Adaptation level as frame of reference for prediction of psychophysical data. *American Journal of Psychology, 60,* 1-29.

Helson, H. (1948). Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychological Review, 55,* 297–313.

Helson, H. (1964). *Adaptation-level theory*. New York, NY: Harper and Row.

Henshaw, E. J., & Freedman-Doan, C. R. (2009). Conceptualizing mental health care utilization using the Health Belief Model. *Clinical Psychology: Science and Practice, 16,* 420–439.

Hohagen, F., Rink, K., Kappler, C., Riemann, D., Weyerer, S., & Berger, M. (1993). Prevalence and treatment of insomnia in general practice. *European Archives of Psychiatry and Clinical Neuroscience, 242,* 329–336.

Holt-Lunstad, J., Smith, T. B., & Layton, B. (2010). Social relationships and mortality: a meta-analysis. *PLoS Medicine, 7,* e1000316.

Horton, J. J., & Chilton, L. B. (2010, June). *The labor economics of paid crowdsourcing*. Paper presented at the 11th Association for Computing Machinery Conference on Electronic Commerce, Cambridge, MA.

House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science, 241,* 540–545.

Hsu, Y. W., Ho, C. H., Wang, J. J., Hsieh, K. Y., Weng, S. F., & Wud, M. P. (2013). Longitudinal trends of the healthcare-seeking prevalence and incidence of insomnia in Taiwan: An 8-year nationally representative study. *Sleep Medicine, 14,* 843–849.

Hyman, H. H. (1942). The Psychology of Status. *Archives of Psychology, 269*, 94-102.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior, 38,* 21-37.

Idler, E. L., & Kasl, S. (1991). Health perceptions and survival: do global evaluations of health status really predict mortality? *Journal of Gerontology, 46*(2), S55–S65.

Idler, E. L., & Kasl, S. (1995). Self-ratings of health – do they also predict changes in functional ability? *Journals of Gerontology Series B – Psychological Sciences and Social Sciences, 50*(6), 344–353.

Idler, E., Russell, L., & Davis, D. (2000). Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study, 1992. *American Journal of Epidemiology, 152*(9), 874-883.

Ingham, I., & Miller, P. (1979). Symptom prevalence and severity in general practice. *Journal of Epidemiology and Community Health, 33,* 191-198.

Jansson- Fröjmark, M., & Linton, S. J. (2007). Psychological mechanisms in the maintenance of insomnia: Arousal, distress, and sleep-related beliefs. *Behaviour Research and Therapy, 45,* 511–521.

Jansson-Fröjmark, M., & Linton, S. J. (2008). The role of psychological mechanisms to insomnia in its early phase: A focus on arousal, distress, and sleep-related beliefs. *Psychology & Health, 23*(6), 691-705.

Johnson, R. C. (2010). The health returns of education policies from preschool to high school and beyond. *The American Economic Review, 100,* 188-194.

Jou, Y. H., & Fukada, H. (2002). Stress, health, and reciprocity and sufficiency of social support: the case of university students in Japan. *Journal of Social Psychology, 142*(3), 353-370.

Jürges, H. (2008). Self-assessed health, reference levels, and mortality. *Applied Economics, 40*(5), 569-582.

Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review, 114,* 678–703.

Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social Science and Medicine, 69*(3), 307-316.

Jylhä, M., Guralnik, J. M., Balfour, J., & Fried, L. P. (2001). Walking difficulty, walking speed, and age as predictors of self-rated health: the Women's Health and Aging Study. *Journals of Gerontology Series A – Biological Sciences and Medical Sciences, 56*(10), M609–M617.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

Kaplan, G. A., & Baron-Epel, O. (2003). What lies behind the subjective evaluation of health status? *Social Science and Medicine, 56,* 1669–1676.

Kaplan, G. A., & Camacho, T. (1983). Perceived health and mortality: a nine-year follow-up of the human Population Laboratory Cohort. *American Journal of Epidemiology, 117,* 292–304.

Kaplan, G. A., Strawbridge, W. J., Camacho, T., & Cohen, R. D. (1993). Factors associated with change in physical functioning in the elderly: a six-year prospective study. *Journal of Aging and Health, 5,* 40–53.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.

Katon, W. J. (2003). Clinical and health services relationships between major depression, depressive symptoms, and general medical illness. *Biological Psychiatry, 54,* 216–226.

Katon, W. J., Russo, J. E., Von Korff, M., Lin, E. H., Ludman, E., & Ciechanowski, P. S. (2008). Long-term effects on medical costs of improving depression outcomes in patients with depression and diabetes. *Diabetes Care, 31,* 1155–1159.

Katon, W., & Ciechanowski, P. (2002). Impact of major depression on chronic medical illness. *Journal of Psychosomatic Research, 53,* 859–863.

Kelstrup, A. M., Juillerat, P., & Korzenik, J. (2014). The accuracy of self-reported medical history: A preliminary analysis of the promise of internet-based research in Inflammatory Bowel Diseases. *Journal of Crohn's and Colitis, 8,* 349–356.

Kerkhofs, M., & Lindeboom, M. (1995). Subjective health measures and state dependent reporting errors. *Health Economics, 4*(3), 221-235.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., … Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32,* 959–976.

Kessler, R. C., Berglund, P. A., Coulouvrat, C., Hajak, G., Roth, T., Shahly, V., … Walsh, J. K. (2011). Insomnia and the performance of US workers: results from the America Insomnia Survey. *Sleep, 34*(9), 1161-1171.

Kessler, R. C., Demler, O., Frank, R. G., Olfson, M., Pincus, H. A., Walters, E. E., … Zaslavsky, A. M. (2005). Prevalence and treatment of mental disorders, 1990 to 2003. *New England Journal of Medicine, 352,* 2515–2523.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis, 9,* 137-163.

Kjvinen, P., Halonen, P., Eronen, M., & Nissinen, A. (1998). Self-rated health, physician-rated health and associated factors among elderly men: the Finnish cohorts of the Seven Countries Study. *Age and Ageing, 27,* 41 -47.

Klein, W. M. (1997). Objective standards are not enough: Affective, self-evaluative, and behavioral responses to social cognition information. *Journal of Personality and Social Psychology, 72,* 763-774.

Knutson, K. L. (2010). Sleep duration and cardiometabolic risk: A review of the epidemiologic evidence. *Best Practice & Research: Clinical Endocrinology & Metabolism, 24,* 731–743.

Krause, N. M., & Jay, G. M. (1994). What do global self-rated health items measure? *Medical Care, 32*(9), 930–942.

Kriegsman, D. M., Penninx, B. W., van Eijk, J. T., Boeke, J. P., & Deeg, D. J. (1996). Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly. *Journal of Clinical Epidemiology, 49,* 1407-1417.

Kroenke, K., & Mangelsdorff, A. D. (1989). Common symptoms in ambulatory care: incidence, evaluation, therapy, and outcome. *American Journal of Medicine, 86*(3), 262-266.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric Annals, 9,* 1–7.

Kroenke, K., Spitzer, R. L., & Williams, J. B., (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606–613.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics, 50*(6), 613-621.

Kroenke, K., Spitzer, R. L., Williams, J. B., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine, 146,* 317–25.

Krueger, J., & Clement, R. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology, 67,* 596–610.

Lack, L. C. (2007). The general public's concept of normal sleep. *Sleep and Biological Rhythms, 5,* A155.

Lakey, B., & Cohen, S. (2000). Social support theory and measurement. In S. Cohen, L. G. Underwood & B. J. Gottlieb (Eds.), *Social support measurement and intervention: A guide for health and social scientists* (pp. 29–52). New York, NY: Oxford University Press.

Layes, A., Asada, Y., & Kepart, G. (2012). Whiners and deniers - what does self-rated health measure? *Social Science and Medicine, 75*(1), 1-9.

Léger, D., Morin, C. M., Uchiyama, M., Hakimi, Z., Cure, S., & Walsh, J. K. (2012). Chronic insomnia, quality-of-life, and utility scores: Comparison with good sleepers in a cross-sectional international survey. *Sleep Medicine, 13,* 43–51.

Leventhal, E. A., Hansell, S., Diefenback, M., Leventhal, M., & Glass, D. C. (1996). Negative affect and self-report of physical symptoms: Two longitudinal studies of older adults. *Health Psychology, 15,* 192-199.

Leventhal, H., Meyer, D., & Nerenz, D. R. (1980). *The common-sense model of illness danger*. In S. Rachman (Ed.), Medical psychology (Vol. II). New York, NY: Pergamon Press.

Lichstein, K. L., Durrence, H. H., Taylor, D. J., Bush, A. J., & Riedel, B. W. (2003). Quantitative criteria for insomnia. *Behaviour Research and Therapy, 41,* 427-445.

Lindeboom, M., & van Doorslaer, E. (2004). Cut-point shift and index shift in self-reported health. *Journal of Health Economics, 23,* 1083–1099.

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology, 73,* 91-103.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. *Medical Care, 46,* 266–274.

Major, B. (1994). From social inequality to personal entitlement: The role of social comparison, legitimacy appraisals, and group membership. In M. P. Zanna (Ed.), *Advances in experimental social psychology Vol. 26* (pp. 293–355). Orlando, FL: Academic Press.

Maltby, J., Wood, A. M., Vlaev, I., Taylor, M. G., & Brown, G. D. A. (2012). Contextual effects on the perceived health benefits of exercise: The exercise rank hypothesis. *Journal of Sport & Exercise Psychology, 34,* 828–841.

Manderbacka, K. (1998). Examining what self-rated health question is understood to mean by respondents. *Scandinavian Journal of Public Health, 26,* 145-153.

Mansson, N., & Rastam, L. (2001). Self-rated health as a predictor of disability pension and death: A prospective study of middle-aged men. *Scandinavian Journal of Public Health, 29,* 151-158.

Marcus, D. K., Gurley, J. R., Marchi, M. M., & Bauer, C. (2007). Cognitive and perceptual variables in hypochondriasis and health anxiety: A systematic review. *Clinical Psychology Review,* 27, 127–139.

Marge, M., Banerjee, S., & Rudnicky, A. I. (2010, March). *Using the Amazon Mechanical Turk for transcription of spoken language.* Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 5270–5273.

Markman, K. D., & McMullen, M. N. (2003). A reflection and evaluation model of comparative thinking. *Personality and Social Psychology Review, 7,* 244–267.

Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *General Hospital Psychiatry, 28*(1), 71-77.

Martin, R., Rothrock, N., Leventhal, H., & Leventhal, E. (2003). Common sense models of illness: Implications for symptom perception and health-related behaviours. In J. Sulls & K. A. Wallston (Eds) *Social Psychological Foundations of Health and Illness* (pp199-225). Malden, MA: Blackwell Publishing.

Mason, W., & Watts, D. J. (2009, June). *Financial incentives and the performance of crowds.* Proceedings of the Human Computation Workshop: Knowledge Discovery and Data Mining Conference. Paris, France.

Mavaddat, N., Kinmonth, A. L., Sanderson, S., Surtees, P., Bingham, S., & Khaw, K. T. (2011). What determines Self-Rated Health (SRH)? A cross-sectional study of SF-36 health domains in the EPIC-Norfolk cohort. *Journal of Epidemiology and Community Health, 65,* 800-806.

McAlaney, J., & McMahon, J. (2007). Normative beliefs, misperceptions and heavy episodic drinking in a British student sample. *Journal of Studies on Alcohol and Drugs, 68*(3)*,* 385–392.

McAteer, A., Elliott, A. M., & Hannaford, P. C. (2011). Describing the size of the symptom iceberg in a UK-wide community-based survey. *British Journal of General Practice, 61,* 12-17.

Meer, J., Miller, D. L., & Rosen, H. S. (2003). Exploring the health-wealth nexus. *Journal of Health Economics, 22,* 713-730.

Meerlo, P., Sgoifo, A., & Suchecki, D. (2008). Restricted and disrupted sleep: Effects on autonomic function, neuroendocrine stress systems and stress responsivity. *Sleep Medicine Reviews, 12,* 197–210.

Mellner, C., & Lundberg, U. (2003). Self- and physician-rated general health in relation to symptoms and diseases among women. *Psychology, Health and Medicine, 8*(2), 123-134.

Melrose, K. L., Brown, G. D. A., & Wood, A.M. (2013). Am I abnormal? Relative rank and social norm effects in judgments of anxiety and depression symptoms severity. *Journal of Behavioral Decision Making, 26,* 174–184.

Menard, S. (1995). *Applied Logistic Regression Analysis.* Thousand Oaks, CA: Sage.

Menard, S. (2004). Six approaches to calculating standardized logistic regression coefficients. *The American Statistician,* 58(3), 218-223.

Menard, S. (2011). Standards for standardized logistic regression coefficients. *Social Forces,* 89(4), 1409-1428.

Merton, R. K., & Kitt, A. (1950). *Contributions to the theory of reference group behavior.* Glencoe, IL: Free Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for information processing. *Psychological Review, 63,* 81–97.

Mojtabai, R. (2008). Social comparison of distress and mental health help-seeking in the US general population. *Social Science and Medicine, 67,* 1944–1950.

Mojtabai, R., Olfson, M., & Mechanic, D. (2002). Perceived need and help-seeking in adults with mood, anxiety, or substance use disorders. *Archives of General Psychiatry, 59,* 77–84.

Morgenthaler, T., Kramer, M., Alessi, C., Friedman, L., Boehlecke, B., Brown, T., … American Academy of Sleep Medicine. (2006). Practice parameters for the psychological and behavioral treatment of insomnia: An update. An American Academy of Sleep Medicine report. *Sleep, 29,* 1415–1419.

Morin, C. M. (1993). *Insomnia: Psychological assessment and management.* New York, NY: Guilford Press.

Morin, C. M., & Jarrin, D. C. (2013). Insomnia and healthcare-seeking behaviors: impact of case definitions, comorbidity, sociodemographic, and cultural factors. *Sleep Medicine, 14,* 808–809.

Morin, C. M., & Ware, J. C. (1996). Sleep and psychopathology. *Applied Preventive Psychology, 5,* 211-224.

Morin, C. M., Blais, F., & Savard, J. (2002). Are changes in beliefs and attitudes about sleep related to sleep improvements in the treatment of insomnia? *Behaviour Research and Therapy, 40,* 741-752.

Morin, C. M., Colecchi, C., Stone, J., Sood, R., & Brink, D. (1999). Behavioral and pharmacological therapies for late-life insomnia: a randomized controlled trial. *JAMA, 281*(11), 991-999.

Morin, C. M., LeBlanc, M., Daley, M., Gregoire, J. P., & Merette, C. (2006). Epidemiology of insomnia: prevalence, self-help treatments, consultations, and determinants of help-seeking behaviors. *Sleep Medicine, 7,* 123–130.

Morin, C. M., Stone, J., Trinkle, D., Mercer, J., & Remsberg, S. (1993). Dysfunctional beliefs and attitudes about sleep among older adults with and without insomnia complaints. *Psychology and Ageing, 8,* 463– 467.

Morin, C. M., Vallières, A., & Ivers, H. (2007). Dysfunctional Beliefs and Attitudes About Sleep (DBAS): validation of a brief version (DBAS-16). *Sleep, 30,* 1547–1554.

Morse, S., & Gergen, K. J. (1970). Social comparison, self-consistency, and the concept of self. *Journal of Personality and Social Psychology, 16,* 148-156.

Muggah, E., Graves, E., Bennett, C., & Manuel, D. G. (2013). Ascertainment of chronic diseases using population health data: a comparison of health administrative data and patient self-report. *BMC Public Health,* 13:16.

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110,* 472–489.

Mussweiler, T., & Epstude, K. (2009). Relatively fast! Efficiency advantages of comparative information processing. *Journal of Experimental Psychology: General, 138,* 1-21.

Mussweiler, T., & Posten, A. (2012). Relatively certain! Comparative thinking reduces uncertainty. *Cognition, 122,* 236-240.

Mussweiler, T., Rüter, K., & Epstude, K. (2004). The ups and downs of social comparison: Mechanisms of assimilation and contrast. *Journal of Personality and Social Psychology, 87*, 832-844.

Niedrich, R. W., Sharma, S., & Wedell, D. H. (2001). Reference price and price perceptions: A comparison of alternative models. *Journal of Consumer Research, 28,* 339-354.

Nielsen, L. S., Danielsen, K. V., & Sorensen, T. I. (2010). Short sleep duration as a possible cause of obesity: Critical analysis of the epidemiological evidence. *Obesity Reviews, 12*(2), 78–92.

Nomura, H., John, P. C., & Cotterill, S. (2011). The use of feedback to enhance environmental outcomes: a randomised controlled trial of a food waste scheme. *Local Environment, 16*(7), 637-653.

Nyawata, I. D., & Topping, A. E. (2006). Symptom interpretation as a source of delay in malignant melanoma presentation—A literature review. *Clinical Effectiveness in Nursing, 9,* e68–e78.

Ohayon, M. M. (2002). Epidemiology of insomnia: what we know and what we still need to learn. *Sleep Medicine Review, 6,* 97–111.

Ohayon, M. M. (2005). Relationship between chronic painful physical condition and insomnia. *Journal of Psychiatric Research, 39*(2), 151-159.

Ohayon, M. M., Carskadon, M. A., Guilleminault, C., & Vitiello, M. V. (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep, 27,* 1255–1273.

Okajima, I., Komada, Y., & Inoue, Y. (2011). A meta-analysis on the treatment effectiveness of cognitive behavioral therapy for primary insomnia. *Sleep and Biological Rhythms, 9,* 24–34.

Okajima, I., Komada, Y., Nomura, T., Nakashima, K., & Inoue, Y. (2012). Insomnia as a risk for depression: a longitudinal epidemiologic study on a Japanese rural cohort. *Journal of Clinical Psychiatry, 73,* 377–383.

Okura, Y., Urban, L. H., Mahoney, D. W., Jacobsen, S. J., & Rodeheffer, R. J. (2004). Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *Journal of Clinical Epidemiology, 57,* 1096–1103.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments using Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5), 411-419.

Parducci, A. (1963). The range-frequency compromise in judgement. *Psychological Monographs, 77,* 565.

Parducci, A. (1965). Category judgement: A range-frequency model. *Psychological Review, 72,* 407-418.

Parducci, A. (1968). The relativism of absolute judgement. *Scientific American, 219,* 84-90.

Parducci, A. (1995). *Happiness, pleasure, and judgement: The contextual theory and its applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Parducci, A., & Perrett, L. F. (1971). Category rating scales: Effects of relative spacing and frequency. *Journal of Experimental Psychology Monograph, 89,* 427-452.

Perkins, H. W. (2007). Misperceptions of peer drinking norms in Canada: another look at the “reign of error” and its consequences among college students. *Addictive Behaviors, 32*(11), 2645–2656.

Perkins, H. W., & Wechsler, H. (1996). Variation in perceived college drinking norms and its impact on alcohol abuse: a nationwide study. *Journal of Drug Issues, 26*(4), 961–974.

Peters, S., Stanley, I., Rose, M., & Salmon, P. (1998). Patients with medically unexplained symptoms: Sources of patients’ authority and implications for demands on medical care. *Social Science and Medicine, 46,* 559-565.

Pinquart, M. (2001). Correlates of subjective health in older adults: a meta-analysis. *Psychology and Aging, 16,* 414-426.

Powdthavee, N. (2009). Ill-health as a household norm: Evidence from other people's health problems. *Social Science and Medicine, 68,* 251-259.

Powdthavee, N., & van den Berg, B. (2011). Putting different price tags on the same health problem: Re-evaluating the well-being valuation. *Journal of Health Economics, 30*(5), 1032-1043.

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioural experiments. *Journal of Theoretical Biology, 299,* 172–179.

Reynolds, J. S., & Perrin, N. A. (2004). Mismatches in social support and psychosocial adjustment to breast cancer. *Health Psychology, 23*(4), 425-430.

Richards, M. A., Westcombe, A. M., Love, S. B., Littlejohns, P., & Ramirez, A. J. (1999). Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet, 353*(9159), 1119-1126.

Rickwood, D. J., Deane, F. P., Wilson, C. J., & Ciarrochi, J. (2005). Young people's help-seeking for mental health problems. *Australian E-Journal for the Advancement of Mental Health, 4,* 218–251.

Riskey, D. R., Parducci, A., & Beauchamp, G. K. (1979). Effects of context in judgements of sweetness and pleasantness. *Perception and Psychophysics, 26,* 171-176.

Ristvedt, S. L., & Trinkaus, K. M. (2005). Psychological factors related to delay in consultation for cancer symptoms. *Psychooncology, 14,* 339–350.

Robine, J-M., Jagger, C., & The Euro-REVES 2 Group. (2003). Creating a coherent set of indicators to monitor health across Europe. *European Journal of Public Health, 13*(Suppl 1), 6-14.

Roecklein, K. A., Carney, C. E., Wong, P. M., Steiner, J. L., Hasler, B. P., & Franzen, P. L. (2013). The role of beliefs and attitudes about sleep in seasonal and nonseasonal mood disorder, and nondepressed controls. *Journal of Affective Disorders, 150,* 466–473.

Rogers, R. W. (1983). Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. In J. T. Cacioppo & R. E. Petty (Eds.), *Social psychophysiology: A source book*. (pp. 153–176). New York, NY: Guilford Press.

Runciman, W. G. (1966). *Relative deprivation and social justice.* London, UK: Routledge and Kegan Paul.

Rusell, J. A., & Fehr, B. (1987). Relativity in the perception of emotions in facial expressions. *Journal of Experimental Psychology: General, 116,* 223-237.

Safer, M. A., Tharps, Q. J., Jackson, T. C., & Leventhal, H. (1979). Determinants of three stages of delay in seeking care at a medical clinic. *Medical Care, 17,* 11–29.

Sánchez-Ortuño, M. M., & Edinger J. D. (2010). A penny for your thoughts: Patterns of sleep-related beliefs, insomnia symptoms and treatment outcome. *Behaviour Research and Therapy, 48,* 125–133.

Sanders, G. S. (1981). The interactive effect of social comparison and objective information on the decision to see a doctor. *Journal of Applied Social Psychology, 11*(5), 390–400.

Sandler, I. N., & Barrera, M., Jr. (1984). Toward a multimethod approach to assessing the effects of social support. *American Journal of Community Psychology, 12*(1), 37-52.

Sarason, B. R., Sarason, I. G., & Gurung, R. A. R. (1997). Close personal relationships and health outcomes: A key to the role of social support. In S. Duck (Ed.), *Handbook of personal relationships* (2nd ed., pp. 547-573). New York, NY: Wiley.

Sarason, B. R., Sarason, I. G., & Pierce, G. R. (1990). Traditional views of social support and their impact on assessment. In B. R. Sarason, I. G. Sarason, & G. R. Pierce (Eds.), *Social support: An interactional view* (pp. 9–25). New York, NY: Wiley.

Sarris, V. (1967). Two critical experiments on Helson's weighted-average model. *The American Journal of Psychology, 80*(3), 331-344.

Sarris, V. (1971). *Perception and judgement.* Gottingen, Germany: Hogrefe.

Scambler, G., & Scambler, A. (1985). The illness iceberg and aspects of consulting behaviour. In R. Fitzpatrick & J. Hinton (Eds.), *The experience of illness* (pp. 32-50). London, UK: Tavistock.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69,* 379–399.

Schultz, P. W. (1999). Changing behaviour with normative feedback interventions: a field experiment on kerbside recycling. *Basic and Applied Psychology, 21*(1), 25–36.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive and reconstructive power of social norms. *Psychological Science, 18*(5)*,* 429–434.

Schüz, B., Wurm, S., Schöllgen, I., & Tesch-Römer, C. (2011). What do people include when they self-rate their health? Differential associations according to health status in community-dwelling older adults. *Quality of Life Research, 20*(10), 1573-1580.

Scott, S., & Walter, F. (2010). Studying help-seeking for symptoms: The challenges of methods and models. *Social and Personality Psychology Compass, 4*(8), 531–547,

Scott, S. E., Walter, F. M., Webster, A., Sutton, S., & Emery, J. (2013). The model of pathways to treatment: conceptualization and integration with existing theory. *British Journal of Health Psychology, 18,* 45–65.

Sedlmeier, P., & Betsch, T. (2002). *Frequency processing and cognition*. Oxford, England: Oxford University Press.

Seta, J. J. (1982). The impact of comparison processes on coactors' task performance. *Journal of Personality and Social Psychology, 42,* 281-291.

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science, 1,* 213–220.

Shaw, C., Brittain, K., Tansey, R., & Williams, K. (2008). How people decide to seek healthcare: A qualitative study. *International Journal of Nursing Studies, 45,* 1516–1524.

Shochat, T., Umphress, J., Israel , A. G., & Ancoli-Israel, S.(1999). Insomnia in primary care patients. *Sleep, 22*(Suppl 2), 359–365.

Silles, M. A. (2009). The causal effect of education on health: Evidence from the United Kingdom. *Economics of Education Review, 28,* 122-128.

Singh-Manoux, A., Martikainen, P., Ferrie, J., Zins, M., Marmot, M., & Goldberg, M. (2006). What does self rated health measure? Results from the British Whitehall II and French Gazel cohort studies. *Journal of Epidemiology and Community Health, 60*(4), 364-372.

Smith, J. P. (1999). Healthy bodies and thick wallets: the dual relation between health and economic status. *The Journal of Economic Perspectives, 13,* 145-166.

Smith, K. V. & Goldman, N. J. (2011). Measuring health status: Self, interviewer, and physician reports of overall health. *Journal of Aging Health, 23*(2), 242–266.

Smith, R. H., Diener, E., & Wedell, D. H. (1989). Intrapersonal and social comparison determinants of happiness: A range-frequency analysis. *Journal of Personality and Social Psychology, 56,* 317-325.

Snead, C. M. (2007). Health, Self-Rated. In G. Ritzer's (Ed.) *Blackwell encyclopedia of sociology.* Blackwell Reference Online.

Speilman, A. J., Caruso, L. S., & Glovinsky, P. B. (1987). A behavioral perspective on insomnia treatment. *The Psychiatric Clinical of North America, 10*(4), 541-553.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166,* 1092–1097.

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review, 11,* 881–911.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology, 53,* 1-26.

Stoller, E. P., (1997). Medical self care. Lay management of symptoms by elderly people. In M. G. Ory and G. DeFries (Eds), *Self-care in later life: Research, program and policy issues* (pp. 24-61). New York, NY: Springer.

Stuart, A. L., Pasco, J. A., Jacka, F. N., Brennan, S. L., Berk, M., & Williams, L. J. (2014). Comparison of self-report and structured clinical interview in the identification of depression. *Comprehensive Psychiatry, 55,* 866–869.

Suls, J. (1986). Notes on the occasion of social comparison theory's thirtieth birthday. *Personality and Social Psychology Bulletin, 12,* 289–296.

Suls, J. (2011). Social comparison processes: Implications for physical health. In H. S. Friedman (Ed.), *The Oxford handbook of health psychology* (online edition). Oxford, UK: Oxford University Press.

Suls, J., & Wheeler, L. (2000). A selective history of classic social comparison theory and neo social comparison theory. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 3-22). New York, NY: Kluwer Academic/Plenum Press.

Suls, J., Gastorf, J. W., & LawhoRn, J. (1978). Social comparison choices for evaluating a sex- and age-related ability. *Personality and Social Psychology Bulletin, 4,* 102-105.

Suls, J., Marco, C. A., & Tobin, S. (1991). The role of temporal comparison, social comparison, and direct appraisal in the elderly self-evaluations of health. *Journal of Applied Social Psychology, 21*(14), 1125–1144.

Suls, J., Martin, R., & Leventhal, H. (1997). Social comparison, lay referral, and the decision to seek medical care. In B. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being.* Mahwah, NJ: Lawrence Erlbaum.

Taylor, S. E., & Lobel, M. (1989). Social comparison activity under threat: Downward evaluation and upward contacts. *Psychological Review, 96,* 569-575.

Taylor, S. E., Neter, E., & Wayment, H. A. (1995). Self-evaluation processes*. Personality and Social Psychology Bulletin, 21,* 1278– 1287.

Thoits, P. A. (1986). Social support as coping assistance. *Journal of Consulting and Clinical Psychology, 54,* 416–423.

Thoits, P. A. (2011). Mechanisms linking social ties and support to physical and mental health. *Journal of Health and Social Behavior, 52,* 145- 161.

Ubel, P. A., Jankovic, A., Smith, D., Langa, K. M., & Fagerlin, A. (2005). What is perfect health to an 85-year-old?: Evidence for scale recalibration in subjective health ratings. *Medical Care, 43*(10), 1054-1057.

Uchino, B. N. (2004). *Social support and physical health: Understanding the health consequences of relationships.* New Haven, CT: Yale University Press.

Uchino, B. N. (2009). Understanding the links between social support and physical health: A life-span perspective with emphasis on the separability of perceived and received support. *Perspectives on Psychological Science, 4,* 236-255.

Uchino, B. N., Bowen, K., Carlisle, M., & Birmingham, W. (2012). Psychological pathways linking social support to health outcomes: A visit with the “ghosts” of research past, present, and future. *Social Science and Medicine, 74,* 949-957.

Undén, A-L. & Elofsson, S. (2001). Health from the patient's point of view. How does it relate to the physician's judgement? *Family Practice, 18*(2), 174-180.

Walter, F. M., Webster, A., Scott, S. E., & Emery, J. (2011). The Andersen model of total patient delay: A systematic review of its application in cancer diagnosis. *Journal of Health Services Research & Policy, 17*(2), 110-8.

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C. (2005). Twelve-month use of mental health services in the United States: Results from the National Comorbidity Survey replication. *Archives of General Psychiatry, 62,* 629–640.

Ware, J. E., & Gandek, B. (1998). Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *Journal of Clinical Epidemiology, 51*(11), 903–912.

Ware, J. E., Kosinski, M., Bjorner, J. B., Turner-Bowker, D. M., Gandek, B., & Maruish, M. E. (2008). *SF-36v2 Health Survey: Administration guide for clinical trial investigators*. Lincoln, RI: QualityMetric Incorporated.

Ware, J. E., Kosinski, M., & Keller, S. D. (1994). *SF-36 physical and mental health summary scales: A user's manual.* Boston, MA: The Health Institute.

Wedell, D. H., & Parducci, A. (1988). The category effect in social judgement: Experimental ratings of happiness. *Journal of Personality and Social Psychology, 55,* 341-356.

Wedell, D. H., Parducci, A., & Geiselmen, R. E. (1987). A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology, 23,* 230-249.

Wedell, D. H., Parducci, A., & Lane, M. (1990). Reducing the dependence of clinical judgment on the immediate context: Effects of number of categories and type of anchors. *Journal of Personality and Social Psychology, 58,* 319-329.

Wedell, D. H., Santoyo, E. M., & Pettibone, J. C. (2005). The thick and the thin of it: contextual effects in body perception. *Basic and Applied Social Psychology, 279,* 213-228.

Weinstein, N. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39,* 806–820.

Wheeler, L. (1966). Motivation as a determinant of upward comparison. *Journal of Experimental Social Psychology, 2*(Suppl. 1), 27-31.

Whitaker, K. L., Scott, S. E., & Wardle, J. (2015). Applying symptom appraisal models to understand sociodemographic differences in responses to possible cancer symptoms: a research agenda. *British Journal of Cancer, 112*(Suppl 1), S27-S34.

White, K. S., Raffa, S. D., Jakle, K. R., Stoddard, J. A., Barlow, D. H., Brown, T. A., … Gervino, E. V. (2008). Morbidity of DSM-IV Axis I disorders in patients with noncardiac chest pain: Psychiatric morbidity linked with increased pain and health care utilization. *Journal of Consulting and Clinical Psychology, 76,* 422–430.

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin, 90,* 245-271.

Wills, T. A., & Gibbons, F. X. (2009). Commentary: Using psychological theory in help-seeking research. *Clinical Psychology: Science and Practice, 16,* 440-444.

Wolff, J. K., Schmiedek, F., Brose, A., & Lindenberger, U. (2013). Physical and emotional well-being and the balance of needed and received emotional support: Age differences in a daily diary study. *Social Science and Medicine, 91,* 67-75.

Wood, A. M., Brown, G. D. A., & Maltby, J. (2012). Social norm influences on evaluations of the risks associated with alcohol consumption: Applying the rank based decision by sampling model to health judgments. *Alcohol and Alcoholism, 47,* 57–62.

Wood, A. M., Boyce, C. J., Moore, S. C., & Brown, G. D. A. (2012). An evolutionary based social rank explanation of why low income predicts mental distress: A 17 year cohort study of 30,000 people. *Journal of Affective Disorders, 136,* 882–888.

Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin, 106,* 231-248.

World Health Organisation, Statistics Netherland. (1996). *Health interview surveys: Towards international harmonization of methods and instruments.* Copenhagen, Denmark: WHO Regional Office for Europe (WHO Regional Publications, European Series, no 58).

Zanna, M. P., Goethals, G. R., & Hill, J. F.(1975). Evaluating a sex-related ability: Social comparison with similar others and standard setters. *Journal of Experimental Social Psychology, 11,* 86-93.

van der Zee, K. I., Buunk, B. P., & Sanderman, R. (1995). Social comparison as a mediator between health problems and subjective health evaluations. *British Journal of Social Psychology, 34,* 53-65.

Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment, 52,* 30-41.

Zimet, G. D., Powell, S. S., Farley, G. K., Werkman, S., & Berkoff, K. A. (1990). Psychometric characteristics of the Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment, 55,* 610–617.

Zimmerman, M., Coryell, W., Wilson, S., & Corenthal, C. (1986). Evaluation of symptoms of major depressive disorder: Self-report vs. clinician ratings. *Journal of Nervous & Mental Disease, 174*(3), 150-153.

Zizi, F., Jean-Louis, G., Brown, C. D., Ogedegbe, G., Boutin-Foster, C., & McFarlane, S. I. (2010). Sleep duration and the risk of diabetes mellitus: Epidemiologic evidence and pathophysiologic insights. *Current Diabetes Reports, 10,* 43–47.

Appendices

Appendix 1: The Patient Health Questionnaire-9 (PHQ-9)

Over the last 2 weeks, how often have you been bothered by the following problems?

Little interest or pleasure in doing things
Not at all □ Several days □ More than half the days □ Nearly every day □

Feeling down, depressed, or hopeless
Not at all □ Several days □ More than half the days □ Nearly every day □

Trouble falling or staying asleep, or sleeping too much
Not at all □ Several days □ More than half the days □ Nearly every day □

Feeling tired or having little energy
Not at all □ Several days □ More than half the days □ Nearly every day □

Poor appetite or overeating
Not at all □ Several days □ More than half the days □ Nearly every day □

Feeling bad about yourself —or that you are a failure or have let yourself or your family down
Not at all □ Several days □ More than half the days □ Nearly every day □

Trouble concentrating on things, such as reading the newspaper or watching television
Not at all □ Several days □ More than half the days □ Nearly every day □

Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual
Not at all □ Several days □ More than half the days □ Nearly every day □

Thoughts that you would be better off dead or of hurting yourself in some way
Not at all □ Several days □ More than half the days □ Nearly every day □

(Kroenke & Spitzer, 2002)

Appendix 2: The Generalised Anxiety Disorder-7 (GAD-7)

Over the last 2 weeks, how often have you been bothered by the following problems?

Feeling nervous, anxious or on edge
Not at all □ Several days □ More than half the days □ Nearly every day □

Not being able to stop or control worrying
Not at all □ Several days □ More than half the days □ Nearly every day □

Worrying too much about different things
Not at all □ Several days □ More than half the days □ Nearly every day □

Trouble relaxing
Not at all □ Several days □ More than half the days □ Nearly every day □

Being so restless that it is hard to sit still
Not at all □ Several days □ More than half the days □ Nearly every day □

Becoming easily annoyed or irritable
Not at all □ Several days □ More than half the days □ Nearly every day □

Feeling afraid as if something awful might happen
Not at all □ Several days □ More than half the days □ Nearly every day □

(Spitzer et al., 2006)

Appendix 3: The Multidimensional Scale of Perceived Social Support (MSPSS)

To what extent do you agree with the following statements:
There is a special person around when I am in need
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

There is a special person with whom I can share my joys and sorrows
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

My family really tries to help me
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

I get the emotional help and support I need from my family
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

I have a special person who is a real source of comfort to me
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

My friends really try to help me
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

I can count on my friends when things go wrong
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

I can talk about my problems with my family
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

I have friends with whom I can share my joys and sorrows
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

There is a special person in my life who cares about my feelings
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

My family is willing to help me make decisions
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

I can talk about my problems with my friends
Very Strongly Disagree □ Strongly Disagree □ Mildly Disagree □ Neutral □ Mildly Agree □ Strongly Agree □ Very Strongly Agree □

(Zimet et al., 1988)

Appendix 4: The Arizona Social Support Interview Schedule – Satisfaction Subscale (ASSIS-S)

How would you rate your satisfaction or dissatisfaction with the times you talked to people about your personal and private feelings during the past month?
Very dissatisfied □ Moderately dissatisfied □ Slightly dissatisfied □ Neither satisfied or dissatisfied □ Slightly satisfied □ Moderately satisfied □ Very satisfied □

During the past month, how satisfied or dissatisfied were you with the things that people loaned or gave to you?
Very dissatisfied □ Moderately dissatisfied □ Slightly dissatisfied □ Neither satisfied or dissatisfied □ Slightly satisfied □ Moderately satisfied □ Very satisfied □

During the past month, how satisfied or dissatisfied were you with advice that you were given?
Very dissatisfied □ Moderately dissatisfied □ Slightly dissatisfied □ Neither satisfied or dissatisfied □ Slightly satisfied □ Moderately satisfied □ Very satisfied □

During the past month, how satisfied or dissatisfied were you with the times that people told you that they liked your ideas or the things that you did?
Very dissatisfied □ Moderately dissatisfied □ Slightly dissatisfied □ Neither satisfied or dissatisfied □ Slightly satisfied □ Moderately satisfied □ Very satisfied □

During the past month, how satisfied or dissatisfied were you with the help you received in doing things that you needed to do?
Very dissatisfied □ Moderately dissatisfied □ Slightly dissatisfied □ Neither satisfied or dissatisfied □ Slightly satisfied □ Moderately satisfied □ Very satisfied □

During the past month how satisfied or dissatisfied were you with the times that you got together with people just to have fun and relax?
Very dissatisfied □ Moderately dissatisfied □ Slightly dissatisfied □ Neither satisfied or dissatisfied □ Slightly satisfied □ Moderately satisfied □ Very satisfied □

(Barrera, 1981)

Appendix 5. The Short Form-36v2 Health Survey (SF-36v2)

For each of the following questions, please select the one box that best describes your answer.

In general, would you say your health is:
Excellent □ Very good □ Good □ Fair □ Poor □

Compared to one year ago, how would you rate your health in general now?
Much better now than one year ago □
Somewhat better now than one year ago □
About the same as one year ago □
Somewhat worse now than one year ago □
Much worse now than one year ago □

The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports
Yes, limited a lot □ Yes, limited a little □ No, not limited at all □

Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf
Yes, limited a lot □ Yes, limited a little □ No, not limited at all □

Lifting or carrying groceries
Yes, limited a lot □ Yes, limited a little □ No, not limited at all □

Climbing several flights of stairs
Yes, limited a lot □ Yes, limited a little □ No, not limited at all □

Climbing one flight of stairs
Yes, limited a lot □ Yes, limited a little □ No, not limited at all □

Bending, kneeling, or stooping
Yes, limited a lot □ Yes, limited a little □ No, not limited at all □

Walking more than a mile
Yes, limited a lot □ Yes, limited a little □ No, not limited at all □

Walking several hundred yards
Yes, limited a lot □ Yes, limited a little □ No, not limited at all □

Walking one hundred yards
Yes, limited a lot □ Yes, limited a little □ No, not limited at all □

Bathing or dressing yourself
Yes, limited a lot □ Yes, limited a little □ No, not limited at all □

During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

Cut down on the amount of time you spent on work or other activities
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Accomplished less than you would like
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Were limited in the kind of work or other activities
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Had difficulty performing the work or other activities (for example, it took extra effort)
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

Cut down on the amount of time you spent on work or other activities
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Accomplished less than you would like
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Did work or other activities less carefully than usual
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?
Not at all □ Slightly □ Moderately □ Quite a bit □ Extremely □

How much bodily pain have you had during the past 4 weeks?
None □ Very mild □ Mild □ Moderate □ Severe □ Very severe □

During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?
Not at all □ A little bit □ Moderately □ Quite a bit □ Extremely □

These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks…

Did you feel full of life?
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Have you been very nervous?
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Have you felt so down in the dumps that nothing could cheer you up?
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Have you felt calm and peaceful?
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Did you have a lot of energy?
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Have you felt downhearted and low?
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Did you feel worn out?
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Have you been happy?
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

Did you feel tired?
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?
All of the time □ Most of the time □ Some of the time □ A little of the time □ None of the time □

How TRUE or FALSE is each of the following statements for you?

I seem to get ill more easily than other people
Definitely true □ Mostly true □ Don't know □ Mostly false □ Definitely false □

I am as healthy as anybody I know
Definitely true □ Mostly true □ Don't know □ Mostly false □ Definitely false □

I expect my health to get worse
Definitely true □ Mostly true □ Don't know □ Mostly false □ Definitely false □

My health is excellent
Definitely true □ Mostly true □ Don't know □ Mostly false □ Definitely false □

(Ware et al., 2008)